Teachers College

Contributions to Education

Columbia University

HV1737
M232

M.C. MIGEL LIBRARY
AMERICAN PRINTING
HOUSE FOR THE BLIND

C.1

THE READABILITY OF CERTAIN TYPE SIZES AND FORMS IN SIGHT-SAVING CLASSES

By HAROLD J. McNALLY, Ph.D.

TEACHERS COLLEGE, COLUMBIA UNIVERSITY
CONTRIBUTIONS TO EDUCATION, NO. 883

Published with the Approval of Professors Arthur I. Gates and Robert L. Thorndike, Co-sponsors

Bureau of Publications
TEACHERS COLLEGE, COLUMBIA UNIVERSITY
NEW YORK 1943

COPYRIGHT, 1943, BY HAROLD J. MCNALLY

HV1737
M232
Copy one

TYPOGRAPHY, PRINTING, AND BINDING IN THE U. S. A. BY
KINGSPORT PRESS, INC., KINGSPORT, TENNESSEE

ACKNOWLEDGMENTS

THE prosecution of a study such as this necessitates the assistance and cooperation of a great many persons. Among the many to whom I am deeply indebted, particular thanks are due:

Dr. Arthur I. Gates, co-sponsor, who will understand that my gratitude exceeds my vocabulary;

Dr. Robert L. Thorndike, co-sponsor, who gave friendly encouragement and invaluable assistance at all stages of the research;

Dr. Harold W. McCormick, Director of the Survey of the Study of the Care and Education of the Physically Handicapped in the City of New York, who made necessary administrative arrangements, and gave valuable advice;

The members of the Subcommittee on Readability of the Committee for the Study of the Care and Education of Physically Handicapped Children in the Public Schools of the City of New York, for valuable criticism and suggestions;

The Mergenthaler Linotype Company, for the loan of essential matrices and other materials;

The A. B. Dick Company, manufacturers of Mimeograph Products, for supplying the stencils and doing the Mimeography necessary in the experiment;

And others, too numerous to mention here, who were of real assistance in the conduct of the study.

H. J. MCN.

TABLE OF CONTENTS

THE READABILITY OF CERTAIN TYPE SIZES AND FORMS IN SIGHT-SAVING CLASSES

I

THE PROBLEM: ITS ORIGIN AND DEFINITION

IMPORTANCE OF SIGHT-SAVING CLASSES

An education suited to the needs and limitations of atypical children is an obligation of a democratic society. Prominent in numbers among such children are those with only partial vision. In 1931, the White House Conference stated that a reasonable estimate of the number of such children in the United States needing special educational provision was approximately one in every 500, or something over 50,000 children.[1] Many of these are already in sight-saving classes, of which there are now 635 in the United States. These classes represent 227 cities, 31 states, the District of Columbia, and the territory of Hawaii.[2] In New York City alone there are 88 schools with sight-saving classes,[3] enrolling approximately 2,000 pupils. In other words, this is a sizable and important undertaking. Any attempt to determine experimentally the best methods or materials for use in these classes is of real practical value.

NEED FOR OBJECTIVE EVIDENCE ON THE READABILITY OF TYPE SIZES IN SIGHT-SAVING CLASSES

One of the problems encountered in providing a sight-con-

[1] White House Conference, *Special Education: The Handicapped and the Gifted,* p. 115. The Century Co., New York, 1931.

[2] These figures were supplied by Mrs. Winifred Hathaway, Associate Director of the National Society for the Prevention of Blindness, in a letter dated January 19, 1942.

[3] Committee for the Study of the Care and Education of Physically Handicapped Children in the Public Schools of the City of New York, *Report of the Reading Subcommittee for the Visually Handicapped,* p. 3. New York City Board of Education, March, 1940.

serving program of education for partially sighted pupils is that of recommending a size of type which is a practical optimum for them with respect to readability. This study was designed to attack that problem by attempting to evaluate the readability of certain types and type sizes for children with handicapped vision in the sight-saving classes of New York City.

The need for studies of this kind has been widely recognized among those directly responsible for the education of partially sighted pupils. At present, the reading materials used in sight-saving classes are commonly printed in extra-large type. Generally this is "Clear Type," a widely spaced and leaded Caslon Bold type face, approximately 6.25 mm. in height for capitals and lower case letters with ascenders or descenders. Not everything the pupils read is available in this type size, however, and teachers commonly meet this lack by enlarging reading matter with the use of special typewriters with large type. The capitals and full height letters on these typewriters range from 5 mm. to 5.5 mm. in height; the lower case letters are from 3.5 mm. to about 4 mm. in height. These heights are uniform on any given machine, of course, varying with the make of the typewriter. In some classes, stencils are cut on these typewriters, and are used to duplicate reading matter with a machine duplicator. In addition, many of the teachers spend large amounts of time reproducing reading matter in large lettering by hand.

A search for objective evidence to support these practices, however, revealed practically none whatever. Studies of the readability of type sizes with reference to persons of normal or near-normal vision are reported, but almost nothing appears with reference to any of the aspects of readability for the partially sighted. Indeed, the only published study of this kind that this investigator found was one by Luckiesh and Moss, and that was concerned primarily with the static factor of *visibility* of 18-point and 24-point type.[4]

[4] Matthew Luckiesh and Frank Moss, *The Science of Seeing*, pp. 266–267. D. Van Nostrand Co., New York, 1937.

The basis for the present use of 24-point type in these classes was provided by a series of studies made by Robert B. Irwin in 1920.[5] These investigations were carried out with pupils in several cities, chiefly in Cleveland, and included types ranging from 10-point to 36-point in size. The study upon which most of the ultimate conclusions were based used 44 words "chosen with special reference to letters such as o, e, a, m, n, b, d, which might readily be mistaken by a child with poor sight, if the type was illegible to him; or to such combinations as lo, li, bi, fi, which must be seen clearly to be read correctly."[6] The criteria were time, errors, average reading distance, and apparent effort.

Irwin concluded that, when other factors such as leading and letter spacing are optimum, "a 24-point type is preferred by most (visually handicapped) children."[7] In an attempt to ascertain the effect of eye condition upon type preference, the data were analyzed according to several visual defects, and the difference in preferences analyzed. The conclusion was that, "The same type . . . which is most legible for children in one group is also most legible for the majority in the other groups."[8]

Other more recent investigators and authorities, including many of those supervising or teaching in these classes, have raised the question of whether 24-point type is the best type for *all* partially-sighted children. In reporting a study of the visibility of 18-point and 24-point types (type face unspecified), Luckiesh and Moss make the following statement:

> For the cases involving medium myopia, . . . subjects obtained visibility values nearly as high as did the group of adults possessing normal vision. Thus for them the 24-point type was probably unnecessarily large. . . . On the other hand, the pupils with cataracts obtained a visibility of only 2.60 and 4.31 when observing 18-point and 24-point type respectively. . . . In other words, 18-point and 24-point types are no more visible to the pupils with cataracts than are 6-point and 9-point types for adults with normal vision. Hence,

[5] White House Conference, *op. cit.*, p. 144.

[6] Robert B. Irwin, "Large Type Reading Tests," p. 4. (Unpublished report of the studies discussed.) American Foundation for the Blind, Inc., New York.

[7] *Ibid.*, p. 10. [8] *Ibid.*

it follows that these pupils need the 24-point type in order to obtain a reasonable degree of visibility.[9]

The report of a subcommittee of the Committee for the Study of the Care and Education of Physically Handicapped in the Public Schools of the City of New York states that:

> Present research reports that deal with written materials, lighting, and the physical conditions of reading for partially-sighted pupils are controversial. Further studies should be made to determine standards for these pupils. . . . These studies should investigate size, kind, color, and ink-density of type, spacing of letters, words and lines, lighting, and other physical conditions important in reading. The relative readability of materials produced by mimeograph, hectograph, typewriter, photo-offset, multigraph, and print, should be considered, including both teacher-made and commercial materials. . . . Recommendations for desirable materials and conditions of reading in these classes in the future depend upon the results of such studies.[10]

The Subcommittee on the Blind and Partially Seeing made the following statement in their section of the White House Conference report:

> If further investigation should show that for many sight-saving class pupils 18-point type would be quite as satisfactory as 24-point, it would be of great help in solving the problem of having a sufficient number of books, since many modern libraries are now specializing in "books for tired eyes." The demand for these from a much wider group of people than the limited number in sight-saving classes would appreciably lower the cost of material, and remove one of the chief objections to the establishment of sight-saving classes.[11]

It seems evident, therefore, that there is a felt need for experimentation to gather more evidence on the readability of type sizes in such classes. Such studies might support any one of the following possibilities:

1. Present usage (24-point type and enlarged typed and duplicated material) makes for the greatest readability with sight-saving pupils.

[9] Luckiesh and Moss, *op. cit.*, pp. 266–267.

[10] Committee for the Study of the Care and Education of Physically Handicapped Children in the Public Schools of the City of New York, *op. cit.*, p. 56.

[11] White House Conference, *op. cit.*, p. 144.

2. Some type (or types) smaller than 24-point is *more* readable for these pupils.
3. Some type smaller than 24-point is *equally* readable for these pupils.
4. Type sizes vary in their readability for these pupils, depending upon the kind of visual aberration.
5. 24-point type is too small for these pupils.
6. Typewritten and/or machine-duplicated materials (of the varieties investigated) are as readable for the partially sighted as is printed material, or at least are acceptable for use with such pupils.
7. Typewritten and/or machine-duplicated materials (of the varieties investigated) are not acceptable for use with the partially sighted from the standpoint of readability.

SELECTION OF TYPES TO BE STUDIED

Type size

With such a range of possible outcomes, and such a variety of printing conditions to be studied, the immediate concern in the formulation of the problem was the selection of type sizes, type style, and method of reproduction of reading matter (printing, typing, Mimeography, etc.) to be investigated. Aside from the unlikelihood of their superiority, type sizes larger than 24-point are rendered quite impractical because of their printing costs. Furthermore, some of the evidence points to the possibility that type sizes *smaller* than 24-point might be acceptable—if not preferable—for many partially sighted children.[12] For these reasons, 24-point type was decided upon as the largest type to be studied.

Other reasons led to the decision to use 12-point type as the lower limit of type size to be studied. Most important in this decision was the unwillingness of the authorities in charge of the classes in which the study was to be made to allow the pupils to read type smaller than this. Then, too, research studies of the readability of type sizes for persons with normal vision seem to

[12] Luckiesh, *op. cit.*, pp. 266–267.

indicate that the type size which is optimum for such persons is 10-point type.[13] If this is true, it appears reasonable to assume that 10-point type is too small for optimum readability in the case of children with handicapped vision.

To provide for intermediate gradations between these two rather arbitrary extremes of 12-point and 24-point type, it was decided to test the readability of 14-point and 18-point type as well. In the event of significant differences in the readability of the type sizes studied, this would provide some basis for the determination of the type size which is optimum from the standpoint of the increments in readability as weighed against the practicalities of printing expense. In other words, if it were to be found that 24-point is the more readable, but that it is only very slightly more so than 18-point, the large savings effected by printing in 18-point type might well outweigh the slight advantage in readability held by the larger type.

Style of type

As this was to be a study of the readability of type *sizes*, it was desirable, of course, to keep the *style* of type as uniform as possible. Furthermore, it was deemed wise in this study to approximate as closely as possible the actual conditions in sight-saving classrooms. As has been mentioned, the type face commonly used in these classes is "Clear Type." This is practically the same as Linotype Caslon No. 3 and Monotype Caslon Bold No. 79. The 12-point and 14-point type sizes, therefore, were set in Linotype Caslon No. 3 (a Caslon Bold). It was originally intended to set the 18-point and 24-point sizes in linotype also, but the prac-

[13] Everett Alderman, "The Effects of Size of Type on Speed of Reading, and the Determination of Various Factors That May Influence the Results," *Pittsburgh Schools*, 13, No. 2, 1938. B. R. Buckingham, "New Data on the Typography of Textbooks," *The Textbook in American Education, Thirtieth Yearbook of the National Society for the Study of Education*, Part II, pp. 93–125. Public School Publishing Co., Bloomington, Ill., 1931. M. Luckiesh and F. Moss, "A Summary of Researches Involving Blink-Rate as a Criterion of Ease of Seeing," *Illuminating Engineering*, 35:19–32, 1940. M. Tinker and D. Paterson, "Studies of Typographical Factors Influencing Speed of Reading. II: Size of Type," *Journal of Applied Psychology*, 13:120–130, 1929.

tical difficulties encountered made this impossible. Consequently, these latter two sizes were set in Monotype Caslon Bold No. 79. These two types (Linotype Caslon No. 3 and Monotype Caslon Bold No. 79) are identical to the ordinary observer and only one skilled in distinguishing between type styles could discern their points of difference.

Typewriter type and Mimeography

To approximate further the actual conditions in sight-saving classes, it was decided to include both enlarged typewritten and machine-duplicated materials. The machine used for typing was a special model built by the Royal Typewriter Company, equipped with a type bearing the trade name of Ampli-Type. This type is 1.25 picas (5 mm.) in height for capitals and for lower case letters with ascenders and descenders. For the lower case letters without ascenders or descenders, it is .875 picas (3.5 mm.) in height. All strokes have a uniform thickness of approximately .5 mm. This machine was used to type test materials and to cut stencils for Mimeographing [14] test materials for the study. Its use was decided upon because machines of this and similar makes are used in sight-saving classes to reproduce in large typewriter type materials that otherwise would be considered too small for the pupils in those classes to read.

Thus, the type faces decided upon for investigation were:

1. 12-point Caslon No. 3, Linotype
2. 14-point Caslon No. 3, Linotype
3. 18-point Caslon Bold No. 79, Monotype
4. 24-point Caslon Bold No. 79, Monotype
5. Ampli-Type, typed
6. Ampli-Type, Mimeographed

DEFINITION OF READABILITY

Throughout the foregoing discussion, the term "readability" has been used frequently. In the literature of types and type

[14] "Mimeograph" is the trade name for the duplicating products of the A. B. Dick Co., Chicago and New York.

sizes, this term has sometimes been used interchangeably with "visibility" and "legibility." It has also been used to denote that quality which renders reading matter pleasurable, appealing, or easy to read from the standpoint of comprehension. As used in this report, the term "readability" is not synonymous with any of these terms or meanings. It refers, rather, to the extent to which a given type size or form lends itself to being read with absence of visual effort—with least "cost to . . . human beings operating as human seeing machines" [15]—when comprehension is held relatively constant. "Visibility," on the other hand, refers more to the degree of clarity with which an object can be distinguished, and is a function of various conditions which also affect readability, including illumination, size of object, distance, contrast, etc. "Legibility" generally connotes the idea of capability of being recognized, particularly with respect to symbolic language, when readability conditions extraneous to the visual object meet supra-threshold conditions. These two conceptions are less dynamic than the meaning of readability herein employed. The three terms are not separate and distinct, however; they are closely related. Readability implies both adequate visibility and legibility, and is the more inclusive of the three concepts. It is largely a physiological concept, and—as has been mentioned—bears an inverse relationship to visual effort: the greater the visual effort necessary to read a given printed passage, the less readable is that passage considered to be. This is the definition of readability which was employed in this study.

SUMMARY

The study herein reported was undertaken because of the need for more objective evidence on the relative readability of several type sizes, and of enlarged typewritten and Mimeographed materials. Present usage, as a result of some not-too-rigidly controlled studies made in Cleveland in 1920, is usually 24-point Caslon Bold type ("Clear Type"), and materials en-

[15] Luckiesh and Moss, *loc. cit.*, p. 12.

larged in various ways. The need for further evidence has been explicitly recognized by authorities in this field. The types selected for evaluation in this study were 12-, 14-, 18-, and 24-point types, and Ampli-Type, typed and Mimeographed.

For the purpose of this study, "readability" is defined in terms of ease of seeing or "cost to the reader": the greater the cost, the less the readability and vice versa.

II

SELECTION OF THE CRITERIA

DIFFICULTY OF LOCATING SATISFACTORY CRITERIA

To make any study of readability as herein defined, it is first necessary to find some criterion to measure the expenditure of visual effort, or amount of visual fatigue—the "cost to the reader." An examination of the literature, however, failed to locate a criterion for the measurement of these factors which is universally agreed upon as being valid, reliable, and sufficiently sensitive.

Inquiries to other likely sources, such as the Harvard Fatigue Laboratory, and prominent research workers in physiology and ophthalmology, convinced the investigator that no entirely satisfactory method of measuring visual fatigue has yet been developed. One which seems to hold some promise—that of speed of accommodation to near and far vision [1]—not only is controversial,[2] but presented two obstacles. First, the authorities in charge of the classes in which the study was made were unwilling—and justifiably so—to allow the pupils to be subjected to any prolonged, fatiguing visual task such as would be necessary to produce significant changes in speed of accommodation. Second, the measure of such accommodation speed is a technical task, for which the investigator had neither the training nor the equipment.

[1] C. J. Robertson, "Effect of Fatigue on the Adjustment of the Eye to Near and Far Vision," *Archives of Ophthalmology*, 17:859–876, 1937. R. E. Wager, "A Method of Measuring Fatigue of the Eyes," *Journal of Educational Psychology*, 13:561–570, 1922.

[2] J. I. Kurtz, "An Experimental Study of Ocular Fatigue," *American Journal of Optometry*, 15:86–117, 1938.

This does not mean that other criteria have not been used with success. A number of investigators report studies of the various factors of readability, in which they used criteria which are relatively easy to apply, and which have given positive results in many cases.

METHODS USED IN OTHER STUDIES OF TYPOGRAPHICAL FACTORS

Several approaches have been made to the study of typographical factors. In general, there have been three distinct types of investigations: (1) studies of visibility, measured in terms of visual thresholds; (2) studies measuring readability by amount of work output; and (3) studies of readability defined as "ease of seeing" or "cost to the reader." Studies of the first two types date from late in the nineteenth century; studies of the "cost to the reader" are relatively recent, and largely confined to the investigations of Luckiesh and Moss.

Studies of visibility

Prominent among the investigations emphasizing the visibility or legibility (the terms are used quite interchangeably in the literature) of letters are those of Roethlein, Tinker and Paterson, and Luckiesh and Moss, although a number of other investigators report studies of this nature. The studies reported mention several methods of appraising this factor. All of them, however, are concerned with what Luckiesh and Moss term "visual thresholds." These two investigators state that visual thresholds vary with the purpose of the visual task, and list the following threshold levels:

"1. Maximum ease of seeing, or minimum cost of seeing to the human seeing machine.
2. Maximum performance of useful work, regardless of cost to the human seeing machine.
3. Maximum certainty of recognition of critical details.
4. Minimum certainty of recognition of critical details.

5. Maximum certainty of distinguishing presence of object.
6. Minimum certainty of distinguishing presence of object." [3]

Most studies of the visibility or legibility of print are concerned with levels 3 and 4. The studies reported, however, use different methods of appraisal.

Roethlein used the distance method to study the legibility of type faces. The types studied were moved toward the observer from a distance at which they were unrecognizable by him, until they could be recognized.[4] This method was also used by Webster and Tinker,[5] Taylor,[6] Sanford,[7] and Holmes.[8]

Burtt and Basch studied the legibility of the isolated letters of several type faces with the use of Weiss's Focal Variator.[9] This apparatus is a system of lenses so related to one another that an image may be projected on a ground glass screen in any degree of clarity, from complete clearness to unrecognizable blur.[10] The measure of the legibility was the degree of focus at which the projected letter could be recognized.

Still another method which has been used is that of controlled tachistoscopic exposure, wherein the legibility is measured by the minimum length of exposure necessary for recognition.[11]

[3] M. Luckiesh and F. Moss, *The Science of Seeing*, pp. 46–47. D. Van Nostrand Co., New York, 1937.

[4] B. E. Roethlein, "The Relative Legibility of Different Faces of Printing Types," *American Journal of Psychology*, 23:1–36, 1912.

[5] H. M. Webster and M. A. Tinker, "The Influence of Type Face on the Legibility of Print," *Journal of Applied Psychology*, 19:43–52, 1935, and "The Influence of Paper Surface on the Legibility of Print," *Journal of Applied Psychology*, 19:145–147, 1935.

[6] C. D. Taylor, "The Relative Legibility of Black and White Print," *Journal of Educational Psychology*, 25:561–578, 1934.

[7] E. C. Sanford, "The Relative Legibility of Small Letters," *American Journal of Psychology*, 1:402–435, 1888.

[8] G. Holmes, "The Relative Legibility of Black and White Print," *Journal of Applied Psychology*, 15:248–251, 1931.

[9] H. E. Burtt and C. Basch, "The Legibility of Bodoni, Baskerville Roman, and Cheltenham Type Faces," *Journal of Applied Psychology*, 7:237–245, 1923.

[10] A. P. Weiss, "The Focal Variator," *Journal of Experimental Psychology*, 2:106–113, 1917.

[11] R. Miyake, J. W. Dunlap, and E. E. Cureton, "The Comparative Legibility of Black and Colored Numbers on Colored and Black Backgrounds," *Journal of General Psychology*, 3:340–343, 1930.

The most recent method of studying type visibility is that of Luckiesh and Moss. They have developed a Visibility Meter, which they describe as an instrument consisting "essentially of two colorless photographic filters with precise gradients in density, which may be rotated in front of the eyes while looking at an object or performing a visual task. These gradient filters not only reduce the apparent brightness of the visual field due to absorption, but also lower the contrast between the object of regard and its background, due to a scattering of light by the photographic film."[12] The "gradient filters" of this instrument are simply strips of photographic film ranging in density from colorless transparency to virtual opacity. The object being studied is viewed through these filters, with effects similar to those which would be attained by looking through smoked glasses of varying densities. The feature of the instrument is that the degrees of density have been calibrated and standardized in terms of standard illumination intensities and visual test objects.

Using this instrument, it is possible to get either "absolute" visibility ratings in terms of the readings on the instrument, or "relative" visibility ratings, expressed in per cent of visibility when a given condition (a given type size, illumination level, type style, etc.) is taken as standard, or 100 per cent. The readings can be interpreted in terms of illumination intensity necessary to equalize the visibility of test objects with that of the standard, or in terms of changes in size necessary to achieve that end.

The relative effects of many of the variables affecting readability have been studied by these two investigators with the use of the Visibility Meter.

[12] M. Luckiesh and F. Moss, "The Visibility and Readability of Print on White and Tinted Papers," *Sight-Saving Review,* 8:123–134, 1938. See also M. Luckiesh, and F. Moss, "Supra-Threshold Visibility," *Journal of the Optical Society,* 30:62–69, 1940; "Visibility and Ease of Seeing," *Industrial Medicine,* 9:33, 1940; and "Visibility: Its Measurement and Significance in Seeing." *Journal of the Franklin Institute,* 220:431, 1935.

Studies of work output as a measure of readability

Some investigators have attempted a more dynamic approach to the problem than that which is implied in the concept of *visibility*. Many studies are reported wherein readability is determined by "work output," which refers here to the amount of material which can be read in a given time. The foremost exponents of this technique are Tinker and Paterson. These two investigators studied not only the readability of type sizes,[13] but many other factors affecting speed of reading as well. Short speed tests with definite time limits were used, and readability was measured in terms of the amount of material read within these time limits.

Speed of production (i.e., *reading* speed) has been used as a criterion of readability by many other investigators also. Baird,[14] Blackhurst,[15] Buckingham,[16] Gilliland,[17] Greene,[18] Griffing and Franz,[19] and Hovde,[20] report studies in which they utilized it to appraise the effects on readability of a number of typographical conditions.

In general, investigators seem to agree that work output as a criterion of readability is not as sensitive as could be desired. Luckiesh and Moss, the foremost critics of the method, have used it in conjunction with their other procedures, and report

[13] M. A. Tinker and D. G. Paterson, "Studies of Typographical Factors Influencing Speed of Reading: II, Size of Type," *Journal of Applied Psychology*, 13:120–130, 1929.

[14] J. W. Baird, "The Legibility of a Telephone Directory," *Journal of Applied Psychology*, 1:30–37, 1917.

[15] J. H. Blackhurst, *Investigations in the Hygiene of Reading*. Warwick and York, Baltimore, 1927.

[16] B. R. Buckingham, "New Data on the Typography of Textbooks," *The Textbook in American Education, Thirtieth Yearbook of the National Society for the Study of Education*, Part II, pp. 93–125, 1931.

[17] A. R. Gilliland, "The Effect on Reading of Changes in Size of Type," *Elementary School Journal*, 24:138–146, 1923.

[18] E. B. Greene, "The Relative Legibility of Linotyped and Typewritten Material," *Journal of Applied Psychology*, 18:697–704, 1934.

[19] H. Griffing and S. I. Franz, "On the Conditions of Fatigue in Reading," *Psychological Review*, 3:513–530, 1896.

[20] H. F. Hovde, "The Relative Effects of Size of Type, Leading and Context," *Journal of Applied Psychology*, 13:600–629 and 14:63–73, 1929 and 1930.

that they are firmly convinced that it is a poor indicator of ease of seeing. At one point for example, they make the statement that:

> Speed of performing work is a relatively insensitive and sometimes misleading criterion of ease or difficulty of seeing. . . . Even though the ocular muscles become fatigued from long periods of reading, it has been shown that they continue to function effectively at the expense of greater effort on the part of the reader. Thus measurements of the ability of the visual mechanism to perform a visual task or test usually do not reveal significant changes in the quality of their performances even after prolonged exertion. Apparently the brain is a hard taskmaster for the organs of vision, since clear binocular vision is demanded and obtained under extreme conditions.[21]

As their chief objection to reading speed as a criterion is its insensitivity, these investigators do concede that "if a statistically significant difference in the rate of reading different materials is found, it may be safely assumed that the materials are different in readability." [22] It is their contention that, since extreme differences in readability would be necessary to affect speed of reading, some more sensitive criterion is necessary for finer discriminations.

Studies of cost to the reader

The third method of study, that of measuring ease of seeing or "cost to the reader," has been used almost exclusively by Luckiesh and Moss. There is a good reason for this. As was pointed out at the beginning of this chapter, there appears to be no criterion of visual effort or fatigue which authorities agree upon as being both valid and reliable. Among those criteria which have been suggested and applied are: changes in amplitude of accommodation,[23] ergographic records of fatigue of con-

[21] M. Luckiesh and F. Moss, "Visibility and Readability of Print on White and Tinted Papers," *Sight-Saving Review,* 8:123–134, 1938.

[22] M. Luckiesh and F. Moss, "The Readability of Stencil-Duplicated Materials," *Sight-Saving Review,* 9:295–304, 1939.

[23] J. I. Kurtz, "An Experimental Study of Ocular Fatigue," *American Journal of Optometry,* 15:86–117, 1938 and R. E. Wager, "A Method of Measuring Fatigue of the Eyes," *Journal of Educational Psychology,* 13:561–570, 1922.

vergence,[24] changes in size of the pupil,[25] metabolism and pulse rate changes,[26] and rate of involuntary eye-blink.[27] With the exception of eye-blink, none of these has been validated by repeated use in experiments to any appreciable extent. Indeed, the investigations of metabolism and pulse rate, and of amplitude of accommodation seem to indicate that these criteria are not satisfactory.

On the other hand, the rate of involuntary eye-blink, which Luckiesh and Moss strongly endorse as a criterion, is reported by them to be quite sensitive to those changes which axiomatically would demand increased visual effort.[28] Blink rate, as a criterion of visual effort being expended, or of visual fatigue, has been used extensively by these two investigators, although no others appear to have used it to any extent. Working from the hypothesis of Ponder and Kennedy that blink rate is closely related to mental tension, and that the eyelid movements "constitute a kind of relief mechanism whereby nervous energy otherwise unutilized, passes into a highly facilitated path," [29] they developed the theory that the rate of blinking is indicative of visual effort and fatigue. They selected many visual tasks which they considered axiomatically fatiguing, and counted the number of blinks made by subjects performing these and correlative tasks less fatiguing in character. They found blink rate to increase *without fail* with:

[24] M. Luckiesh and F. Moss, "Fatigue of Convergence Induced by Reading as a Function of Illumination Intensity," *American Journal of Ophthalmology,* 18:319–323, 1935.

[25] M. Luckiesh and F. Moss, "Size of Pupil as a Possible Index for Visual Fatigue," *American Journal of Ophthalmology,* 16:393–396, 1933.

[26] R. McFarland, C. Knehr, and C. Berens, "Metabolism and Pulse-Rate as Related to Reading under High and Low Levels of Illumination," *Journal of Experimental Psychology,* 25:65–75, 1939.

[27] M. Luckiesh and F. Moss, "The Eyelid Reflex as a Criterion of Ocular Fatigue," *Journal of Experimental Psychology,* 20:589–596, 1937.

[28] M. Luckiesh and F. Moss, "Criteria of Readability," *Journal of Experimental Psychology,* 27:256–270, 1940.

[29] E. Ponder and W. P. Kennedy, "On the Act of Blinking," *Quarterly Journal of Experimental Physiology,* 18:89–110, 1927–1928.

1. An increase in the duration of the reading period.
2. The introduction of an improper visual correction while reading.
3. The addition of a glare-source within the visual field.
4. A marked decrease in level of illumination on reading matter.
5. A decrease to below normal (assuming "normal" to be about 10-point to 12-point type) in size of type read.
6. A decrease from 6 points in leading of type read.
7. A marked decrease in boldness of type read.
8. The use of background papers of saturated colors.

In a summary of the findings of these investigations, Luckiesh and Moss state that "the criterion of blinking, when used under proper experimental conditions, appears to be an unusually sensitive one."[30] Elsewhere, they state that:

> It is becoming increasingly certain from numerous investigations which we have made during recent years that the relative ease with which different materials are read may be revealed by observing the rate of involuntary blinking while reading under controlled conditions. . . . In addition, there is a wealth of evidence from the sciences of physiology and psychology which reveals the significance of this criterion from a fundamental viewpoint of ocular strain and fatigue. In brief, this criterion of ease of seeing possesses the advantages of fundamental significance, simplicity in application, objectiveness, and satisfactory sensitivity. The consistently successful results obtained in applying this criterion confirm our general theory that ease of seeing is more readily and certainly revealed by involuntary reactions than it is by criteria of performance over which the subject exercises voluntary control. Speed of reading may be cited as an example of the latter.[31]

It is to be noted that these investigators quite consistently speak of blink rate as a criterion of "ease of seeing," and only occasionally as a criterion of ocular fatigue. From their investigations, it seems apparent that they consider this criterion sensi-

[30] M. Luckiesh and F. Moss, "A Summary of Researches Involving Blink-Rate as a Criterion of Ease of Seeing," *Illuminating Engineering*, 35:31. 1940.

[31] M. Luckiesh and F. Moss, "The Readability of Stencil-Duplicated Materials," *Sight-Saving Review*, 9:295–304, 1939.

tive to increases in the expenditure of visual effort, even before an appreciable amount of fatigue is induced. For example, the statement is made in one of their reports, that "the rate of blinking while performing critical visual tasks is a function of both the duration and severity of the specific visual task; and . . . an augmentation of either of these factors increases the rate of blinking." [32] If this is true, one need not demand that the subject perform a task to the exhaustion point, nor even long enough for fatigue to become evident, in order to measure the relative ease with which a given size of type is read. If a change in type size results in an augmentation of the severity of the reading task, it should be reflected directly in an increase in eye-blink rate.

CRITERIA TO BE USED IN THIS STUDY

As Luckiesh and Moss have used this criterion of eye-blink rate in many investigations of seeing ease, described in scores of articles, and report that: "Rate of blinking *invariably* increased as the task was made more difficult in various ways," [33] and as there is no reported evidence disputing its validity, it appeared to be an ideal criterion for this investigation.

Nevertheless, as other investigators report positive results from the use of speed and accuracy of comprehension in investigations of readability, it was decided to use both reading speed and eye-blink rate as criteria of readability in this study.

Furthermore, since Luckiesh and Moss claim that visibility, as measured by their Visibility Meter, correlates well with readability, it was decided to measure with that instrument the visibility of the materials used in this study. The relative visibility of the types studied could then be compared with their readability, as revealed by reading speed and eye-blink rate.

[32] M. Luckiesh and F. Moss, "The Eyelid Reflex as a Criterion of Ocular Fatigue," *Journal of Experimental Psychology*, 20:589–596, 1937.

[33] M. Luckiesh and F. Moss, "A Summary of Researches Involving Blink-Rate as a Criterion of Ease of Seeing," *Illuminating Engineering*, 35:31, 1940. Italics are author's.

SUMMARY

Efforts to determine a criterion of the readability of printed matter led to the discovery that no criterion has been established universally as being valid for this purpose. The search revealed that three approaches have been made to the problem of determining the characteristics of printed matter which make for the optimum in readability: (1) studies of visibility, measured in terms of visual thresholds; (2) studies of readability measured by speed of work output; and (3) studies of readability measured in terms of cost to the reader. Numerous investigators report studies (for persons of normal vision) of the first two types, using a number of devices and methods. In the field of reading, studies of work output have generally been measurements of reading speed.

Studies of readability in terms of cost to the reader have been confined largely to the extensive investigations of Luckiesh and Moss, in which rate of involuntary eyelid reflex has been the criterion. Other criteria have been investigated, some of which have been demonstrated to be definitely undesirable; others have not been validated to any appreciable extent. Since Luckiesh and Moss claim that eye-blink rate invariably increases with an increase in the severity of the visual task, this appeared to be an ideal criterion for the purpose of the present study. Reading speed has also given positive results in a number of studies by different experimenters; consequently it also was included as a criterion. In addition, it was decided to measure visibility with the Luckiesh-Moss Visibility Meter, to make possible a comparison of this factor with the readability revealed by the other criteria.

III

CONDITIONS OF THE EXPERIMENT

LENGTH OF THE SAMPLING PERIODS

Even had it been possible to use a criterion of readability which would be universally agreed upon as being valid and reliable, the question of the length of testing time which is necessary to assure a practical optimum of reliability for the criterion scores would still remain. In this respect there were limiting factors. Here, too, those responsible for administering the education of the partially sighted children being studied were quite understandably unwilling to allow the children to read for more than very brief periods, particularly in reading the types smaller than 18 points in size. Furthermore, it will be remembered that six type variations were investigated. The pattern of the experiment (see page 42) required as many test forms as there were type variations; namely, six. However, it is difficult to find six reading passages of equivalent difficulty, or suitable tests with six equivalent forms, of sufficient length to yield more than a few minutes' sampling of the criteria.

Other investigators, however, have obtained positive results with sampling periods of only a few minutes. Luckiesh and Moss, in particular, state that: "In general, our experiences indicate that a period of five minutes is adequate, although in some cases it is advisable to repeat the series of measurements and use the mean value for interpretation."[1] As was observed on page 16, this would seem to be particularly true in the case of the eye-

[1] M. Luckiesh and F. Moss, "Frequency of Blinking as a Clinical Criterion of Ease of Seeing," *American Journal of Ophthalmology*, 22:616–621, 1939.

blink criterion, which—according to Luckiesh and Moss—directly reflects the relative severity of the visual task.

Tinker and Paterson used even shorter periods in studying readability,[2] using a time span of only 1¾ minutes. It will be recalled that their studies used reading speed as an indicator of readability.

On these bases, therefore, it was decided that each subject should read each type variation for a period of five minutes.

SELECTION OF READING TESTS TO BE USED

Criteria to be met

In selecting reading tests to be used in a study of this kind, there are several criteria to be met. The first is of obvious importance: the technique of the test must be such as to insure that the material is read. This necessitates some kind of objective check on comprehension. Second, it is desirable to have no comprehension difficulties of any consequence in the reading matter of the test materials. This is necessary to insure a rate score with accuracy as a negligible factor, so as to allow the influence of type size to exert its maximum effect. Third, the reliability of the test forms used must be sufficiently high to assure the meaningfulness of the group comparisons. In this respect, Lindquist states that: "A reliability as low as .40 may be adequate for comparisons of average scores for large groups of individuals."[3]

A fourth criterion, peculiar to this experiment, is that the test must provide reading material sufficient to fill a five-minute period, which had been decided upon as the time span of each sample.

For the conditions of this experiment, it was also necessary that the test have six fairly equivalent forms, one for each size of type to be tested.

[2] M. A. Tinker and D. G. Paterson, "Studies of Typographical Factors Influencing Speed of Reading, II: Size of Type," *Journal of Applied Psychology*, 13:120–130, 1929.

[3] E. F. Lindquist, *A First Course in Statistics*, pp. 201–202. Houghton Mifflin Company, Boston, 1938.

The tests selected

After an examination and consideration of a number of kinds of tests, it was decided that the speed of reading tests developed by Gates [4] were the most practical for this study. These tests are made up of short paragraphs of from 25 to 30 words. Following each item are four words, one of which is the correct answer to the question asked or implied in the paragraph. The person being tested is to underline the correct answer. An example of these items is the paragraph below, selected at random from Form A of the tests as they were printed for this experiment:

> 47. The Indians brought the Pilgrims presents of baskets, corn and furs. They helped the Pilgrims and taught them many things they needed to know. What kind of Indians were they?
> sleepy friendly bad warlike

It is evident that the comprehension of each item in the test is checked as it is read, satisfying the first criterion.

The difficulty level of the paragraphs in these tests is approximately third grade. As steps were taken to insure that each subject for the experiment had a reading ability *above* third grade level, the second criterion of the negligibility of comprehension difficulties is met.

Reliability of the tests

It was earlier implied that the reliability coefficient of the test forms used must be at least .40, preferably better. Gates reports the correlations between the two forms of the Reading Survey to be from .87 to .90.[5] These correlations were obtained by computing the correlations of the two forms with each other. For the Modern School Achievement Tests, the coefficients of re-

[4] A. I. Gates, Modern School Achievement Tests, Short Form (Speed of Reading) Forms I and II. Bureau of Publications, Teachers College, Columbia University, New York, 1931 and Gates Reading Survey (Speed Test), Forms I and II. Bureau of Publications, Teachers College, Columbia University, New York, 1939.

[5] A. I. Gates, *Manual of Directions for Gates Reading Survey*, pp. 13–14. Bureau of Publications, Teachers College, Columbia University, New York.

liability are on the order of .87.[6] It can be seen that in this respect the tests are more than satisfactory.

Adaptation of the tests to this study

With respect to the criterion of number of forms (it will be remembered that six were necessary) and test length, these tests seemed to fail to meet the requirements. To begin with, there are but four forms in print, comprising altogether 228 test items. As these items are all closely equivalent in reading difficulty, it was thought possible to redistribute them randomly in such a manner as to make six test forms of 38 items each. Dr. Gates agreed that this was a defensible procedure.

It was not at all certain, however, that the better readers might not finish this number of items before the termination of the five-minute sampling period. Consequently, the investigator wrote seventy-two additional items, identical in nature with those in the published tests, and closely equivalent in difficulty. These items were distributed randomly throughout the six test forms, which then consisted of 50 items each. The investigator wished to verify his own judgment as to the similarity of these "home-made" items to those of Gates's tests. Consequently, he asked several persons practiced in the use, analysis, and construction of reading tests to examine the tests in which these items were scattered, and to try to distinguish the items written by the investigator from those of the original tests. In no case were the newly written items distinguishable from the others, with respect to authorship or difficulty.

Because of the fact that the new items were so similar in nature and reading difficulty to the original test items, it is extremely unlikely that they should materially affect the reliability coefficients of those tests. Particularly is it unlikely that the coefficients would be affected so greatly as to render the tests useless for purposes of group comparison.

[6] *Manual of Directions for Modern School Achievement Tests, Short Form,* p. 1. Bureau of Publications, Teachers College, Columbia University, New York.

The tests finally decided upon, then, were six forms of 50 items each, the items randomly assigned to the forms from the Gates Reading Survey, The Modern School Achievement Test of Reading Speed, and items written by the investigator.

TYPOGRAPHICAL CONDITIONS OF THE TESTS

Conditions of the printed tests

In deciding upon the typographical conditions (other than size and style of type) of the tests, an effort was made to make them consistent with the most enlightened practice. Even for persons with normal vision, however, the optimum for length of line, amount of leading, spacing, etc., is not agreed upon by investigators. Thus it is evident that the specifications for these conditions could not be supported by scientific evidence. This is explained in part by the complex of interrelationships among these factors, and the influence of habit on the part of readers. For example, the optimum for leading (space between lines) and line length varies with the size and style of type. This may be affected, also, by the fact that the reader is used to reading a given type size set in lines of a given length, and with a given leading.[7]

The only alternative was to rely upon expert advice from authoritative sources. Consequently, a number of authorities versed in the most approved practices in typography were consulted. Among them were an executive of a type manufacturing company, an executive in the manufacturing department of a prominent publishing house, a book designer for a well-known book manufacturer, and an instructor for an established printing school.[8] It is believed, therefore, that the conditions of leading, spacing, line length, margins, placement of items, and general

[7] A. I. Gates, "What Do We Know about Optimum Lengths of Lines in Reading?" *Journal of Educational Research,* 23:1–7, 1931.

[8] Those who were particularly helpful in this respect were: Mr. Harry L. Gage, Vice-President, Mergenthaler Linotype Company; Mr. Ernst Reichl, designer, H. Wolff Book Manufacturing Company; Mr. John Benbow, Head of the Manufacturing Department, Longmans, Green and Company; and Mr. Thomas J. O'Connell, Instructor, New York School of Printing.

allover arrangements of the tests are in accordance with best printing practice.

The general conditions with respect to these factors for each printed type size were as follows:

12-point Caslon Bold No. 3, Linotype

1. 24-pica line length (111.5 mm., or 4 in.).
2. Approximately 45 characters to a 24-pica line, right-hand margin unjustified.
3. Identifying numbers flush with left-hand margin.
4. Cast on 16-point body (equivalent of 4 points' leading).
5. In each item, one em between the period after the identifying number and the first character of the item.
6. One blank slug (1.5 picas) between the body of the item and the four response words.
7. Two blank slugs between the response words of one item and the beginning of the next item.
8. Two ems between the response words of an item, and the group of four words centered on the page.
9. Five items to a page (13 blank slugs and 20 to 25 lines of print).

14-point Caslon Bold No. 3, Linotype

1. 30-pica line length (127 mm., or 5 in.).
2. Approximately 50 characters to a 30-pica line, right-hand margin unjustified.
3. Identifying numbers flush with left-hand margin.
4. Cast on 18-point body (equivalent to four points of leading).
5. One blank slug (1.5 picas) between the body of the item and the four response words.
6. Two blank slugs between the response words of one item and the beginning of the next item.
7. Two ems between the response words of an item, and the group of four words centered on the page.

8. Four items to a page (10 blank slugs and 16 to 20 lines of print).

18-point Caslon Bold No. 79, Monotype

1. 36-pica line length (152.5 mm., or 6 in.).
2. Approximately 45 characters to a 36-pica line, right-hand margin unjustified.
3. Identifying numbers flush with the left-hand margin.
4. Six points of leading (equivalent to being cast on 24-point body).
5. Two picas between the body of the item and the four response words.
6. Four picas between the response words of one item and the first line of the next item.
7. Two ems between the response words of an item, and the group of four words centered on the page.
8. Three items to a page.

24-point Caslon Bold No. 79, Monotype

1. 40-pica line length (169 mm., or 6⅝ in.).
2. Approximately 40 characters to a 40-pica line, right-hand margin unjustified.
3. Identifying numbers flush with the left-hand margin.
4. Six points of leading (equivalent to being cast on 30-point body).
5. 5, 6, 7, and 8 same as for the 18-point.

Conditions of the typed and Mimeographed tests

The style and size of type for the typewritten tests have been described in Chapter I. All the typing was done by the same person, an experienced and expert typist. Standards for the work were very high, no errors or erasures being allowed. As soon as the typing showed signs of becoming appreciably lighter, the ribbon was changed. The ribbon used throughout was Vogue Black, No. 52 Bulletin Typewriter Ribbon, sold by the Royal Typewriter Company. Even with these precautions, however,

the typing on some pages was heavier than that on others, as is revealed in the visibility readings in Table 1.

The ratings in Table 1 are relative values. They were arrived at by arbitrarily assigning to the most "visible" type (24-point Caslon) a rating of 100, and expressing the Visibility Meter readings for the other types as *percentages* of the visibility reading for 24-point Caslon.

TABLE 1

RELATIVE VISIBILITY RATINGS OF THE EXPERIMENTAL TYPES (IN TERMS OF 24-POINT CASLON, ARBITRARILY ASSIGNED A RELATIVE VISIBILITY VALUE OF 100)*

Type	*Rating*
12-point Caslon	47
14-point Caslon	58
18-point Caslon	85
24-point Caslon	100
Ampli-Type (typed) light	81
Ampli-Type (typed) dark	91
Ampli-Type (Mimeographed)	71

* Obtained by using the Luckiesh-Moss Visibility Meter, averaging the results of the readings of adult observers of normal and near-normal vision. The readings were obtained in the General Electric Research Laboratories under the direction of Dr. Matthew Luckiesh and Mr. Frank K. Moss.

The machine-duplicated copies were run from Mimeograph stencils, on a Mimeograph machine duplicator, by employees of the A. B. Dick Company, manufacturers of Mimeograph products. The stencils were cut by the typist who prepared the typewritten materials, and the same typewriter was used. Because of the large size of the type, it was found desirable to place a sheet of Pliofilm [9] over the face of the stencil to insure a clear and uniform impression. The stencils were cut very deliberately, and all capitals were struck twice. The finished copies were carefully inspected and passed on by an executive of the company. The resulting product was the best Mimeography

[9] Registered trademark, Goodyear Tire and Rubber Company.

that could be produced by the employees of a company with a tradition of good workmanship to uphold. In uniformity and neatness, it was much superior to the typewritten copies, although it was somewhat lighter than the typing.

The page arrangement of the typed and Mimeographed materials was made to resemble as closely as possible that of the 24-point type. The specifications for these materials were as follows:

1. 6-pica left-hand margin (25 mm., or 1 in.).
2. 39-pica line length (165 mm., or 6.5 in.).
3. 39 characters to a 39-pica line, right-hand margin unjustified.
4. Identifying numbers flush with left-hand margin.
5. 0.75 picas (1/4 in.) between lower end of lower case descenders of one line and upper end of lower case ascenders in line below.
6. 2.5 picas (7/16 in.) between descenders of last line in the body of the item, and ascenders of response words.
7. 5 picas (27/32 in.) between descenders of response words of one item and ascenders of first line of next item.
8. Response words spaced according to length, and centered on the page.
9. Three items to a page.

All the tests were printed on paper 8.5 inches by 11.0 inches.

THE PAPER USED FOR PRINTING THE TESTS

The paper decided upon for use in the tests was chosen for several reasons. An effort was made to get a paper closely similar in color to that of the "Clear Type" publications widely used in sight-saving classrooms. At the same time, it was necessary that the paper be one which would "take" printing, typing, and Mimeographing well. The paper finally decided upon was Warren's Olde Style Wove, India, Substance 60. It is a light cream-buff colored book paper of high quality, with a dull, light-dif-

fusing surface. The characteristics of this paper with respect to glare, opacity, surface smoothness, and color are specified in Table 2. The glare test was run on the Bausch and Lomb glarimeter, the opacity test on the Bausch and Lomb opacimeter, and the smoothness test on the Bekk smoothness tester.

TABLE 2

CERTAIN CHARACTERISTICS OF THE PAPER UPON WHICH TESTS WERE PRINTED

Characteristic	*Rating*
Glare (per cent)	
Wire	6.7
Felt	4.8
Opacity (per cent) (contrast ratio)	92.4
Surface Smoothness	
Wire	20.0
Felt	14.2
I. C. I. Color Specification (Illuminant C)	
L (mmu)	576.5
p	21.8
Y	80.8

The values in Table 2 are interpreted as follows:

Glare: The glare values are the ratio, expressed in per cent, of the amount of light reflected from the paper to that reflected by a plate of polished black glass when the angles of incidence and reflection are both 75 degrees (with respect to the perpendicular). The values given in Table 2 are relatively low.

Opacity: The opacity is a measure of the contrast between areas of the paper covering black and white backgrounds. For example, the value of 92.4 means that an area over a black background reflects 92.4 per cent as much light as an area over a (standardized) white background. In general, this ratio is directly related to opacity.

Surface Smoothness: Bekk Smoothness is the method commonly employed for measuring the smoothness of paper. It is an

indirect method, based on the fact that the rate of leak of air between a flat surface and the paper surface depends upon the smoothness of the latter. The Bekk instrument employs an optically flat annular surface, which is placed against the paper under a standard load. Air under a standard average pressure is then permitted to leak from a reservoir connected with the central space defined by the paper and the hole in the annular ring. The smoothness is expressed in terms of the time required for the air pressure to fall from a definite value above the standard pressure to a second definite value below the standard pressure.

I. C. I. Color Specification: "Illuminant C" is a carefully standardized light source which simulates the light from a lightly overcast sky. "L" is the dominant wave length of the color, expressed in millimicrons (one millimicron = 0.000001 millimeters); "p" is the purity of the color; "Y" is the visual efficiency of the color, as seen under illuminant C. If spectrum light of wave length "L" is mixed with the light of illuminant C in a proportion properly corresponding to "p," and of intensity corresponding to "Y," the color of the specimen as seen by the "standard observer" will be matched. For hues ranging from violet through blue, green, etc., to red, the dominant wave length varies from about 400 to 700 millimicrons. The purity "p" for a pure white or a pure gray is zero, and 100 for a "saturated" color (approached, for example, by a red signal light). The visual efficiency is sometimes called visual brightness, and is a quantity which measures the sensation of lightness and darkness; "Y" for pure black is zero, and 100 for a nearly perfect white (magnesium oxide or magnesium carbonate).

SUBJECTS

Selection of the subjects

The design of the experiment (see Chapter IV) called for 72 subjects. These subjects were chosen from the sight-saving classes of New York City. According to the Inspector of the

Classes for the Blind, the general standard of admission to the New York City sight-saving classes is visual acuity of 20/50 or less in the better eye, with correction. Children whose visual acuity is less than 20/200 in the better eye, wearing correction, are admitted to the Braille classes. Frequently, however, pupils with better than 20/50 vision are admitted to the sight-saving classes when the examining ophthalmologist judges that the children have an eye condition which would be seriously aggravated by the visual tasks required in the regular classrooms. The reason, of course, is that defective visual acuity in distance seeing is only one of the evidences of visual defect.

The teachers of the classes in which the experiment was to be conducted were asked to select pupils whose reading abilities were definitely above third grade level. Because of the fact that there are no reading tests with norms standardized on a comparable population, no objective test scores could be obtained for these pupils. Consequently, the selection of the pupils who satisfied the reading level requirement had to be left to the judgment of the teachers. As the only condition was that the pupils be able to read above third grade level, however, it was felt that such judgment could be relied on.

When this preliminary selection had been made, the visual diagnoses of the pupils chosen were carefully examined. Those pupils who verged on blindness, or who had an eye defect which might be aggravated by the test conditions, were eliminated from consideration. A prominent eye physician assisted with this phase of the selection. From the group of pupils who met the foregoing conditions satisfactorily, 12 groups of six pupils each were selected. These 12 groups were from 12 classes, situated in 11 different schools in the boroughs of Manhattan and the Bronx, New York City.[10]

[10] Miss Frances Moscrip, Inspector of the Classes for the Blind in the Public Schools of the City of New York, gave permission to test the pupils used in the experiment, and gave access to the records of the children tested.

TABLE 3

VISUAL DEFECTS OF THE SUBJECTS OF THE EXPERIMENT, WITH SNELLEN AND NEAR-VISION READINGS

Subject	Visual Defect	*Snellen Reading*		*Near-Point Accommodation Reading*		
		O.D.	*O.S.*	*O.D.*	*O.S.*	*O.U.*
1.	Progressive myopia	20/40	20/30	70/500	60/500	70/500
2.	High myopia	20/30	20/30	700/600	650/600	50/500
3.	Progressive myopia	20/70	20/70	40/600	35/600	30/600
4.	High myopia	15/100	15/100	115/600	60/600	55/500
5.	Albino; hyperopic astigmatism	20/70	20/70	255/800	180/1000	125/800
6.	Myopia	20/200	20/20	30/1000	75/300	60/300
7.	High myopia	20/50	20/50	80/300	75/300	75/300
8.	Compound myopic astigmatism	20/70	20/50	40/300	40/300	35/300
9.	Corneal opacity	20/30	5/200	110/300	80/700	80/300
10.	Myopia—secondary to rickets	20/50	20/200	30/300	40/300	40/300
11.	High myopia	20/70	20/70	85/300	90/300	40/300
12.	High myopia with astigmatism	20/100	20/100	45/300	45/300	45/300
13.	Compound myopic astigmatism	20/70	20/70	80/300	45/300	45/300
14.	Hypermetropia; astigmatism; amblyopia	20/40	20/50	73/400	63/600	73/400
15.	High myopia	20/30	20/50	60/300	65/300	60/300
16.	Partial optic atrophy	20/100	20/50	80/600	50/300	35/300
17.	Myopic astigmatism	20/30	20/50	75/300	75/300	75/300
18.	Nystagmus; albinism	20/200	20/200	55/900	50/900	55/900
19.	High myopia	20/70	20/50	95/300	84/300	82/300
20.	Compound hyperopic astigmatism	20/100	20/50	110/500	*	85/400
21.	High myopia	20/100	20/200	60/300	65/300	60/300

22. Hyperopic astigmatism	20/200	20/70	63/300	65/300	65/300
23. Optic atrophy	20/100	20/100	75/900	70/900	66/900
24. Hyperopic astigmatism	20/50	20/50	110/400	90/200	100/400
25. High myopia	20/70	20/70	80/400	80/400	60/400
26. High myopia	20/200	20/200	†	†	†
27. High myopia	20/100	20/200	20/300	20/300	20/300
28. High myopia	20/100	20/100	50/300	45/300	50/300
29. High myopia	20/70	20/50	40/300	40/300	40/300
30. High myopia	15/20	15/20	40/200	50/200	40/200
31. Hyperopia	20/70	20/20	76/300	75/300	62/300
32. High myopia	20/30	20/30	50/300	62/300	67/300
33. Myopia	20/70	20/100	71/700	65/700	60/700
34. High myopia	20/100	20/100	87/300	80/300	44/300
35. Hyperopic astigmatism	20/70	20/50	105/500	115/500	80/500
36. Congenital cataracts; compound myopic astigmatism	20/100	20/100	70/600	75/500	60/500
37. High myopic astigmatism	20/70	20/50	30/300	32/300	32/300
38. Myopic astigmatism	20/100	20/100	52/300	43/300	42/300
39. Mixed astigmatism	20/40	20/50	55/400	70/400	58/400
40. Myopia; divergent strabismas	20/100	20/40	83/300	85/300	75/300
41. High myopia	20/50	20/50	75/400	120/400	120/400
42. Congenital cataracts	20/100	20/70	75/900	70/900	50/900
43. Myopia	20/200	20/200	72/300	85/300	83/300
44. High progressive myopia	20/40	20/40	30/300	45/300	25/300
45. Myopia; compound myopic astigmatism	20/100	20/100	60/400	40/400	40/400
46. High myopia	20/70	20/70	no vision	75/1000	75/1000
47. Compound myopic astigmatism	20/70	20/70	25/400	35/400	25/400
48. High myopia—poor distant vision	10/200	10/100	45/300	45/300	35/300
49. High myopia	20/100	20/50	30/400	35/400	35/400
50. Myopic astigmatism; nystagmus	20/70	20/70	50/700	95/700	45/700

TABLE 3 (*Continued*)

Subject	Visual Defect	Snellen Reading		Near-Point Accommodation Reading		
		O.D.	*O.S.*	*O.D.*	*O.S.*	*O.U.*
51.	Progressive myopia	20/70	20/70	14/300	60/300	50/300
52.	Nystagmus; compound divergent strabismus	20/200	20/50	130/300	65/300	65/300
53.	High myopia	20/100	20/40	80/300	30/300	30/300
54.	High myopia	20/50	20/70	35/300	35/300	35/300
55.	Nystagmus; esotropia	20/200	20/200	75/1000	85/800	85/800
56.	Compound hyperopic astigmatism	20/70	20/70	70/500	85/500	65/500
57.	High myopia	20/50	20/70	45/300	45/300	30/300
58.	Hyperopic astigmatism	20/70	20/50	80/700	65/300	65/300
59.	High myopia; astigmatism	20/200	20/200	45/300	60/300	45/300
60.	High myopia; astigmatism	20/100	20/50	48/300	45/300	45/300
61.	Myopia	20/70	20/70	95/300	95/300	74/300
62.	High myopia	20/60	20/50	62/300	69/300	56/300
63.	High myopia	20/70	20/70	30/300	25/200	39/300
64.	Congenital retinal degeneration	20/25	20/100	31/300	175/900	32/300
65.	High myopia	20/50	20/50	45/300	65/300	40/300
66.	High myopia	20/50	20/70	50/300	48/300	45/300
67.	High myopia	20/40	20/70	35/300	50/300	40/300
68.	Albino; nystagmus; moderate grade mixed astigmatism	20/100	20/100	50/900	40/900	50/900
69.	High myopia with double vision	20/40	20/40	70/300	80/300	60/300
70.	High myopia; myopic astigmatism	20/40	20/40	180/400	135/300	60/300
71.	Compound myopic astigmatism	20/70	20/40	75/300	80/300	70/300
72.	Compound hyperopic astigmatism; nystagmus	20/70	20/100	50/300	50/300	50/300

* Could not be determined.

† Subject would not cooperate.

Characteristics of the population

The children finally selected as subjects for the experiment ranged from 9 years of age to 14 years of age, inclusive. Approximately 80 per cent of them were about equally distributed through the ages of 10, 11, and 12. Their grade placements ranged from grade 4 to grade 8, with the great majority in grades 5 and 6. Of the total, 29 were boys, and 43 were girls.

Table 3 gives the visual diagnosis for each of the 72 subjects. Some of the diagnoses were made by private physicians; others are diagnoses made at public health eye clinics. The near-point of accommodation readings were made by this examiner, using the card of the Ophthalmological Foundation. Near-point of accommodation is the nearest point to the eyes at which the subject is able to focus without blurring of the smallest type he can read. It is measured in millimeters, and expressed in fractional form, with the millimeter reading as the numerator, and the type size rating as the denominator.

The high incidence of myopia and high myopia among these children is typical of the population of the sight-saving classes of New York City. Indeed, no other single eye defect approaches myopia in its range of occurrence among these pupils.

It had been planned to select the pupils to form several groups according to eye defect, so that comparisons of the readability of types according to visual anomaly could be made. It was found, however, that such grouping was not feasible, and the plan was abandoned in favor of a group chosen without respect to eye defect.

As was mentioned earlier in this chapter, the reading ability of all 72 pupils was above third grade level, according to the judgment of the teachers of the pupils. Also in the judgment of the teachers, the best readers in the group were of high sixth, or low seventh, grade in reading level. The great majority were of about good fifth grade level in reading; that is to say, they could read with facility from books ordinarily used in grade five.

TESTING CONDITIONS AND METHODS

The pupils were tested in groups of six, with no other pupils present in the room. The teacher of the class being tested timed the test with a stop watch. To insure uniformity of lighting, the pupils were seated in such a way that the light fell on the papers from the left, except in the case of left-handed pupils, who were seated with the light to their right. Each desk surface was tested with a General Electric Light Meter (which measures light intensity in foot-candles), and every desk surface was receiving between twenty-seven and thirty-five foot-candles of light while the tests were being given.

An observer was seated to the right (to the left for left-handed pupils) and slightly to the rear of each pupil being tested. The observers code-marked each test, and made certain that the pupils understood the directions for taking the test. During the five-minute sampling period, each observer counted the eye-blinks of his subject, and at the conclusion of the five minutes entered on the test paper of the subject the number of eye-blinks counted.

Before the testing for each group began, the author gave the directions for the test procedure in these words:

> We are going to have some easy tests of reading. One of our visitors is going to sit beside each of you to make sure that you do things in the right way. You are not to ask questions about the *reading*, but our visitors will help you if you do something wrong.
>
> Now let us look at the directions. (Reading the directions which appear on the front of each test.) "This is a test to see how well you can read. In it are a number of paragraphs like the one below. Under each paragraph are four words. You are to read the paragraph, and then draw a line under the one word which is right. Remember, draw a line under only *one* word after each paragraph."
>
> Now you do the sample yourself.
>
> (Pupils do the sample, with assistance from the observer if necessary. When all have finished it satisfactorily, examiner continues.)
>
> When Miss ________ says, "Begin," you are to turn this front page back and start reading. Do all the paragraphs the way you did this sample. As you finish each page, turn to the next, and keep right on

reading until Miss ______ says, "Stop." Read as many paragraphs as you can without making mistakes. If you see that you have made a mistake, do not erase, but cross out the line this way (illustrating on blackboard), and draw another line under the one word you think is correct. Remember to draw a line under only *one* word.

At the end of exactly five minutes, the teacher said, "Stop."

The pupils took all six test forms at one sitting. There were three-minute rest periods between the tests, with the exception of the third and fourth test forms. Between these two there was a five-minute rest period, during which the pupils were free to leave their seats, talk with each other, and go to the lavatory. As will be explained in Chapter IV, all six pupils in each group took the same test form at the same time, but at any given time no two pupils were reading the same type variation.

The tests were scored by two persons, each of whom checked his scoring once, and then checked the scoring of the other. Thus, the scoring of each test was checked twice.

Observers, and method of counting eye-blink

The observers who counted the eye-blinks were six adult graduate students. Each was known to be a careful, thorough worker, well acquainted with experimental technique, and aware of the importance of exactitude in experimentation. Before testing the pupils included in this study, each observer had ample practice in counting the eye-blinks of other persons. The author gave careful instructions to the group concerning the method to be used.

Before beginning the experiment, a mechanical counter was tried in counting blinks. It was found that the noise made by the counter distracted the subjects, and this device was discarded. The observers, therefore, simply counted the blinks mentally. This undoubtedly introduced an element of error into the scores. Luckiesh and Moss use this technique,[11] and it has been mentioned that they report consistently positive results.

[11] According to a letter to the author from Frank K. Moss.

SUMMARY

Five minutes was decided upon as ample time for each of the six sampling periods.

Six test forms of 50 items each were used, composed of items from the Gates Modern School Achievement Test of Reading Speed, the speed test from the Gates Reading Survey, and 72 items closely similar to the items in these tests, written by the author of this study. The entire 300 items were randomly distributed through the six test forms.

The typographical conditions of the tests were determined by advice from professional authorities in the field of typography. It is believed, therefore, that the typography of the tests conforms to the best current typographical practice.

The paper used was Warren's Olde Style Wove, India, Substance 60. It is a light cream-buff colored book paper of high quality, with a dull, light-diffusing surface.

The subjects were 72 pupils in sight-conservation classes of the New York City public school system. They were chosen without respect to eye defect; consequently, the group is heavily weighted with myopic subjects. All subjects were judged by their teachers to be above third grade level in reading ability. There were 12 groups of six pupils each, from 11 different schools in the boroughs of Manhattan and the Bronx, New York City. Their ages ranged from 9 to 14, and they were from grades 4 through 8. Of the total, 29 were boys and 43 were girls.

The pupils were tested in groups of six, under fairly uniform conditions. They each took six tests, each test in a different type variation, each lasting five minutes. During this period, an observer sat beside each pupil, and counted the total number of eye-blinks during each five-minute period. All six forms were given at one sitting to each group, with brief rest periods between tests.

IV

EXPERIMENTAL DESIGN AND STATISTICAL METHOD

IMPORTANCE OF EXPERIMENTAL DESIGN

The relationship between experimental design and valid conclusions has been overlooked all too frequently in educational research. Yet that relationship is of crucial importance. The great majority of educational experiments are concerned with measurable differences between groups, or differences between initial and subsequent measures on the same group. The experimenter is generally interested in determining whether differences or changes revealed by such measures have been effected by some given treatment or factor.

If we could be certain that all factors other than that under study were perfectly controlled, we could then be certain that the changes we observed could be attributed to that factor alone. It is well known, however, that a perfectly controlled experimental situation never exists, particularly in educational studies, in which there are so many human variables. Hence, it is always possible that observed differences between initial and final measures may be attributable to factors other than the one whose effect we are attempting to determine, or that real differences induced by the factor being studied are obscured by the effects of uncontrolled or partially controlled variables. Consequently, we can never treat the results of this kind of research as "absolute" data; they must always be looked upon as approximations of the actual conditions, approximations into which the uncontrolled factors have introduced a certain amount of error.

Fortunately, statistical methods have been evolved to estimate the effects of such uncontrolled variables. The differences caused by these factors are considered to be "errors of measurement," and the statistical estimates of their maximum effects are known as "estimates of error."

With the foregoing in mind, we can see that there are two conditions to be satisfied in any educational experiment. First, steps must be taken to provide adequate control of as many variables as possible; second, provision must be made for arriving at a dependable estimate of the effect of variables not fully controlled—a dependable estimate of error.

If the second condition is met, the fact that some factors whose effects we wish to discount are but imperfectly controlled need give us little concern, provided they have been randomized. Our estimate of error will give us an approximation of their maximum influence upon the results. The importance of such an estimate is well stated by Lindquist:

> It is extremely significant that unless one has a fairly definite and dependable (even though subjectively derived) estimate of the maximum error which might be present in an obtained difference, one can draw no useful conclusion from that difference, no matter how precise it may be in reality. So long as the error may be of any magnitude, it is always conceivable that the difference is due to error alone. Given a dependable estimate of error, one can demonstrate that certain hypotheses are inconsistent with the results obtained; without any such estimate any hypothesis whatever is admissible, including of course the hypothesis that there are no real differences in treatments. In a very real sense, then, it is more important to know the *degree* of precision of an experiment, whether high or low, than it is that the precision be in reality high.[1]

The manner in which an experiment is designed will in large measure determine the accuracy of the error estimate. Some experimental designs tend to exaggerate the effects of extraneous factors, leading us to believe that our observed differences are less significant than they actually are. Other types operate to

[1] E. F. Lindquist, *Statistical Analysis in Educational Research*, p. 77. Houghton Mifflin Company, Boston, 1940.

underestimate the magnitude of error, leading to conclusions that the observed differences are more significant than they are in actuality.[2]

DESIGN OF THE EXPERIMENT

After considering experimental designs possible in the present experiment, it was decided to utilize *analysis of variance*, developed in recent years by R. A. Fisher. This method of statistical analysis has as yet had relatively little application in the field of educational research. The technique it employs provides for a dependable estimate of the error inherent in this kind of experiment.

The fundamental thesis of analysis of variance is that the total variance of a large sample containing a number of groups equal in size can be divided into two components: the average of the variances within the groups, and the variance among the group means. These two variances provide independent estimates of the population parameter, one being an estimate of the population variance based on the average variance within the groups of the sample, the other being an estimate of the population variance based on the variance of the means of the sample groups. Assuming that the samples from which these two estimates were derived are random samples of the same population, these estimates of the population variance should differ only by chance. This hypothesis can be tested by means of the F-test, to determine whether the ratio between these two parts of the total variance is larger than chance allows. If it is, we have evidence that the subgroups are not samples of the same population, and that a true difference exists between them.[3]

One of the conditions of this procedure is that all factors which might introduce systematic differences must be randomized.

[2] For a more thorough treatment of this principle, see Lindquist, *op. cit.*, Chap. IV.

[3] This is a drastic précis and oversimplification of the topic. For a more thorough development of the logic of the proposition, see E. F. Lindquist, *Statistical Analysis in Educational Research*, Chapter V, "Analysis of Variance." Houghton Mifflin Company, Boston, 1940.

The necessity for this precaution can be readily understood from the fact that the F-test is a test of the degree to which the variation among the observed differences is greater than that to be expected by chance, and this can be estimated only when random selection has been provided for.

Pattern of test administration

To randomize the effects of the various factors, the tests were administered in a Latin Square pattern to each group of six pupils. As there were six test forms, each form printed in six type variations, there were altogether thirty-six combinations of test form and type variation. In the first group of six pupils, each pupil took six of these combinations, each pupil taking six combinations different from those taken by any of the other five pupils. The same process was repeated with the other five groups of subjects in the first group of thirty-six children, except that the order of test administration was changed in each case. This resulted in assuring that every possible combination of type size, test form, and order of test administration (216 such combinations were possible) was encountered by each of the two groups of thirty-six pupils, without any two pupils within a group taking the same six combinations. In Figure I, for example, it can be observed that Pupil 1 is the only pupil of that group of thirty-six to take Form C in 12-point type in third position of test order. This entire procedure was then replicated with the second group of thirty-six pupils.

ANALYSIS OF THE DATA

An extension of the basic proposition of analysis of variance allows us to make more intricate analyses. Instead of estimating the contribution (to the total variance) of only *two* of the sources of variation, this technique makes possible an analysis of the variance into a *number* of components in such a way as to estimate the variance induced by each, as well as the amount of variance attributable to undetermined or unrandomized factors

	A	B	C	D	E	F
1.	T	M	12	14	18	24
2.	M	12	14	18	24	T
3.	12	14	18	24	T	M
4.	14	18	24	T	M	12
5.	18	24	T	M	12	14
6.	24	T	M	12	14	18

	B	C	D	E	F	A
7.	T	M	12	14	18	24
8.	M	12	14	18	24	T
9.	12	14	18	24	T	M
10.	14	18	24	T	M	12
11.	18	24	T	M	12	14
12.	24	T	M	12	14	18

	C	D	E	F	A	B
13.	T	M	12	14	18	24
14.	M	12	14	18	24	T
15.	12	14	18	24	T	M
16.	14	18	24	T	M	12
17.	18	24	T	M	12	14
18.	24	T	M	12	14	18

	D	E	F	A	B	C
19.	T	M	12	14	18	24
20.	M	12	14	18	24	T
21.	12	14	18	24	T	M
22.	14	18	24	T	M	12
23.	18	24	T	M	12	14
24.	24	T	M	12	14	18

	E	F	A	B	C	D
25.	T	M	12	14	18	24
26.	M	12	14	18	24	T
27.	12	14	18	24	T	M
28.	14	18	24	T	M	12
29.	18	24	T	M	12	14
30.	24	T	M	12	14	18

	F	A	B	C	D	E
31.	T	M	12	14	18	24
32.	M	12	14	18	24	T
33.	12	14	18	24	T	M
34.	14	18	24	T	M	12
35.	18	24	T	M	12	14
36.	24	T	M	12	14	18

Figure I: PLAN OF TEST ADMINISTRATION; LATIN SQUARE PATTERN

Arabic numerals refer to subjects and capital letters to test form. Entries in cells refer to type variations, as follows: numerals are type size; M is Mimeography; T is Typing.

(these latter lumped together as one factor). The F-test can then be applied to determine the significance of the difference among the variances attributable to these factors. In the present study the data were analyzed to estimate the variance derived from each of five sources:

1. Differences among individuals
2. Order of test administration
3. Test form
4. Type variation
5. Error (residual variance)

In some experiments using the technique of analysis of variance, the error (or residual) variance is further analyzed for the variance attributable to the *interaction* among various factors. In this study, for example, it is possible (though improbable) that some of the variance might have been assignable to the interaction between type variation and order of test administration. That is, a given type form might show better results in readability when administered first, rather than third or last. Even were such interaction probable, however, it would have been impossible, in the pattern of this experiment, to analyze its effects. To do so, it would have been necessary for each subject to experience every possible combination of test form, type variation, and test order—clearly a practical impossibility. Because of this, the effects of such interaction (such as they might be) are confounded with the individual differences variance.

In making an analysis of variance, four quantities must be determined for each component: The sum of squares, the degrees of freedom, the variance, and the F-value. Given below are the formulas by which an estimate of the variance of each of the major sources of variation in this study may be derived. The numerator quantity in each case is the sum of squares for that factor, whereas the denominator is the number of degrees of freedom. The symbolism, as it is used herein, is explained beforehand:

S^2 = variance (factor indicated by subscript)
N = number of scores (432)
n = number of subjects (72)
k = number of groups (6 in every case)
X = any score
X_i = score of any individual
X_o = score of any designated test order position, from 1 to 6
X_f = score of any designated test form, A to F
X_t = score of any designated type variation

Basic formulas

(1) Total variance:

$$S^2_T = \frac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{N}}{df = (N - 1)}$$

(2) Variance induced by differences among individuals:

$$S^2_i = \frac{\Sigma\dfrac{(\Sigma X_i)^2}{k} - \dfrac{(\Sigma X)^2}{N}}{df = (n - 1)}$$

(3) Variance induced by order of test administration:

$$S^2_o = \frac{\Sigma\dfrac{(\Sigma X_o)^2}{n} - \dfrac{(\Sigma X)^2}{N}}{df = (k_o - 1)}$$

(4) Variance induced by test forms:

$$S^2_f = \frac{\Sigma\dfrac{(\Sigma X_f)^2}{n} - \dfrac{(\Sigma X)^2}{N}}{df = (k_f - 1)}$$

(5) Variance induced by type variation:

$$S^2_t = \frac{\dfrac{(\Sigma X_t)^2}{n} - \dfrac{(\Sigma X)^2}{N}}{df = (k_t - 1)}$$

(6) Residual variance:

$$S^2 \text{ resid.} = \frac{\begin{matrix}\text{Total sum} \\ \text{of squares}\end{matrix} - \begin{pmatrix}\text{Sum of squares for all other} \\ \text{systematic forms of variance}\end{pmatrix}}{df = N - (n-1) - (k_o - 1) - (k_f - 1) - (k_t - 1)}$$

Analyses of the above components were made for each of the two criteria: speed and eye-blink. The results of these analyses are given in Tables 4 and 5. As we are trying to determine whether or not the systematic variance for each of the components is significantly greater than the error variance, the F-value for each is computed by using the variance of the component concerned as the numerator of a fraction whose denominator is always the residual variance. The resulting F-value can then be referred to a table of such values computed by Snedecor,[4] which gives the values exceeded by chance in 5 per cent of the cases and in 1 per cent of the cases, for various combinations of degrees of freedom. It will be observed from the tables that the F-values with respect to type variation are not significant for either of the two criteria. In every case, the variances arising from differences in type variation are smaller than the variance attributable to any other source, and are even smaller than the variation to be expected from chance alone. In other words, the statistical analysis revealed no systematic differences among the type variations in the criterion scores. The means for each type

Table 4

ANALYSIS OF VARIANCE FOR SPEED SCORES

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Variance*	*F-Values*	$F_{.05}$	$F_{.01}$
Individual differences	9,203	71	129.62	36.929	1.37	1.56
Test form	448	5	89.60	25.527	4.37	9.04
Testing order	116	5	23.20	6.610	4.37	9.04
Type variation	7	5	1.40	.400	4.37	9.04
Residual (error)	1,211	345	3.51			
Total	10,985	431	25.49			

[4] G. W. Snedecor, *Statistical Methods,* Table 10.2, pp. 174–177. Reproduced in Lindquist, *op. cit.,* Table 4, pp. 62–65.

TABLE 5

ANALYSIS OF VARIANCE FOR EYE-BLINK SCORES

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Variance*	*F-Values*	$F_{.05}$	$F_{.01}$
Individual differences	54,953	71	773.99	22.598	1.37	1.56
Test form	175	5	35.00	1.022	4.37	9.04
Testing order	113	5	22.60	.660	4.37	9.04
Type variation	105	5	21.00	.600	4.37	9.04
Residual (error)	11,816	345	34.25			
Total	67,162	431	155.83			

variation are presented in Table 6, to illustrate how little difference there was among them.

TABLE 6

MEANS OF EYE-BLINK AND SPEED SCORES FOR EACH TYPE VARIATION

	T	*M*	*12*	*14*	*18*	*24*	*General Mean*
Eye-blink	15.50	15.78	16.17	15.71	16.57	15.00	15.79
Speed	15.36	15.21	15.19	15.47	15.10	15.28	15.27

Although the *F*-value with respect to type variation is that with which this study is primarily concerned, it is revealing to note the values for the other components. As is to be expected, the variance attributable to the differences among the subjects was the largest and most highly significant for both criteria. For the speed scores (Table 4, p. 46) the next most significant variance was that arising from differences in test form. There are, of course, no reading tests with several forms which are perfectly equated. The fact that this lack of equivalence is revealed and accounted for in the *F*-test is an illustration of the value of the factorial design employed, and of analysis of variance. It will be noted that the analysis also accounted for the variation traceable to the order in which the tests were taken.

The high degree of significance of the variance arising from the effects of individual differences for both speed and eye-blink scores, and of the variance attributable to test form and testing order in the speed scores, serves to emphasize the necessity for controlling these sources of variation, or for providing for a dependable estimate of their maximum effects. The value of the statistical technique employed in this investigation is evidenced by the fact that it provided an estimate of the effects of the major sources of variation, which would probably have been unaccounted for to a large extent had another technique been employed.

Analysis by visual anomaly

Because it was possible that some true differences were hidden in the data, a further analysis was made. From the 72 subjects were chosen 40 with myopia or myopic astigmatism, and 10 with hyperopia or hyperopic astigmatism. Children with these defects were chosen for two reasons: (1) because these defects appeared with greater frequency than any others among the subjects, and (2) because it was felt that if there were any differences in readability induced by visual anomaly, they would be most clearly exhibited and contrasted in near-sighted and far-sighted subjects.

The procedure in this analysis was that of finding within each group (and for each of the two criteria) the mean for each type variation, and determining whether there were significant differences among or between them. This was done for both speed of reading and eye-blink frequency for each of the groups. The method employed to test the significance of the differences between means was that of comparing each difference with its own standard error. The difference would be considered significant if it were as much as 2.576 times its own standard error (the .01 level).

In making an analysis of this kind with these data, it must be recognized that the scores for the various type forms tested are

correlated. That is, the subjects who read fastest while reading one type form are very likely those who read fastest in the reading of the other forms as well. Similarly, those with a normally high blink rate will be likely to blink more frequently than the other subjects, no matter what type forms they are reading. This need give little concern in treating and interpreting the data, as the problem here is *each individual's* variation from type form to type form, regardless of whether his speed or blink rate be consistently high or low. In other words, we are investigating the differences among the scores for type form, not the differences among individuals. Nevertheless, the correlations induced by this factor must be taken into consideration in the statistical analysis. In the analysis of variance, it was accounted for in the "Individual Differences" variance. In making an analysis by the method of testing the significance of differences between means, this correlation must again be accounted for, as the differences being tested are differences between the means of related measures.

Consequently, the sixty correlations among the scores were computed. They ranged from .670 to .931 for the eye-blink scores, with a median of .790, and from .648 to .958 for the speed scores, with .818 as the median. All were positive. These r's were then employed in the formula:

$$\sigma_{M_x - M_y} = \sqrt{\sigma_{M_x} + \sigma_{M_y} - 2r_{xy}\sigma_{M_x}\sigma_{M_y}}$$

to compute the standard error of each difference between means.[5]

For the myopes, the standard error of the mean was arrived at by using the formula: [6]

$$\sigma_M = \frac{S_{(\text{sample})}}{N}.$$

[5] E. F. Lindquist, *A First Course in Statistics*, pp. 124–126. Houghton Mifflin Company, Boston, 1938.
[6] *Ibid.*, pp. 105–115.

This formula was not appropriate for estimating the standard error of the mean for the hyperopes, however, as the sample was so small (10 cases). Consequently, the formula:

$$\sigma_M = \frac{S_{(\text{sample})}}{N - 1}$$

was employed, as it gives a more dependable estimate for small samples (less than 25 cases).[7]

The means of the scores for these two groups are presented in Table 7. Table 8 gives the results of the foregoing analysis. It will be noted that, of the sixty differences between means which were tested, only three were significant at the .01 level. For at least two reasons, these probably may be disregarded. First, out of the sixty differences, it is not unlikely that there will be a few significant ones simply by chance. Second, the differences which are significant do not "make sense." For the myopes, both significant differences are between "adjacent" sizes of type. Whereas the differences between 14- and 18-point and 18- and 24-point type are significant, the difference between 14- and 24-point type is not. For hyperopes, the difference be-

TABLE 7

MEANS OF SPEED AND EYE-BLINK SCORES FOR MYOPES AND HYPEROPES, WITH DIFFERENCES

	Type Size					
	T	*M*	*12*	*14*	*18*	*24*
Speed						
Myopes	15.43	15.65	15.78	15.85	15.65	15.50
Hyperopes	13.00	12.20	12.60	13.20	12.00	12.70
Difference	2.43	3.45	3.18	2.65	3.65	2.80
Eye-Blink						
Myopes	15.08	15.68	15.90	13.90	16.55	14.25
Hyperopes	13.00	15.00	16.50	17.00	15.80	13.10
Difference	2.08	.68	−.60	−3.10	.75	1.15

[7] *Ibid.*, p. 113, footnote.

TABLE 8

DIFFERENCES BETWEEN MEANS, WITH STANDARD ERRORS ($\sigma_{Diff.}$), FOR SPEED AND EYE-BLINK SCORES OF MYOPES AND HYPEROPES

Difference between Means of	Eye-Blink				Speed			
	Myopes		*Hyperopes*		*Myopes*		*Hyperopes*	
	Diff.	$\sigma_{Diff.}$	*Diff.*	$\sigma_{Diff.}$	*Diff.*	$\sigma_{Diff.}$	*Diff.*	$\sigma_{Diff.}$
T and M	.60	1.16	2.00	1.74	.22	.42	.80	.52
T and 12	.82	1.68	3.50	2.44	.35	.54	.40	.49
T and 14	1.18	1.57	4.00	2.28	.42	.46	.20	.99
T and 18	1.47	1.42	2.80	2.52	.22	.46	1.00	1.04
T and 24	.83	1.39	.10	1.46	.07	.41	.30	1.00
M and 12	.22	1.38	1.50	2.24	.13	.51	.40	.65
M and 14........	1.78	1.46	2.00	2.10	.20	.47	1.00	.95
M and 18	.87	1.37	.80	2.80	.00	.51	.20	1.02
M and 24	1.43	1.04	1.90	1.63	.15	.58	.50	.94
12 and 14	2.00	1.51	.50	2.79	.07	.42	.60	1.45
12 and 18	.65	1.42	.70	1.81	.13	.52	.60	.82
12 and 24	1.65	1.11	3.40	1.98	.28	.57	.10	.99
14 and 18........	2.65 *	.99	1.20	2.02	.20	.37	1.20	1.00
14 and 24	.35	1.61	3.90 *	1.55	.35	.54	.50	1.33
18 and 24	2.30 *	.72	2.70	1.67	.15	.44	.70	1.25

* Exceeds 2.576 times the standard error.

tween 14- and 24-point type is significant, but not that between 12- and 24-point type, nor that between 12- and 18-point type. We are probably safe in assuming, therefore, that no real differences among the type variations are indicated by this analysis.

Whereas the foregoing analysis yielded no results justifying any conclusion that one type is more readable than another for either myopes or hyperopes, some differences between the mean scores for the two groups are suggestive. It will be noted in Table 7, for example, that there is a consistent difference between the means of the speed scores for the myopes and hyperopes. For the eye-blink scores, however, the differences are not so consistent. Here the blink rate of the myopes exceeded that of the hyperopes *for all but the smallest type sizes.* This difference was greatest for typing (which approximated the 24-

point type in size) and 24-point Caslon, the largest of the type sizes. For the smallest type sizes, the hyperopes blinked more frequently than the myopes, the greatest difference occurring in the case of 14-point type.

Relationship between the criteria

As one of the purposes of the study was to appraise the relationship obtaining between the criteria of reading speed and eye-blink frequency, the data were examined for what they could reveal in that respect. For each subject, the highest and lowest reading speed scores were ascertained, and compared with the corresponding eye-blink scores. If the criteria were related, we should expect the great majority of cases to be consistent in showing either a direct or an inverse relationship. Actually, a direct relationship was revealed in 32 cases, an inverse relationship in 34 cases, with 6 cases showing no relationship. In the experimental group, therefore, the criteria showed little or no relationship.

SUMMARY

The experiment was set up for a five-component analysis of variance. The tests were administered in a Latin Square pattern, wherein each child took all six forms and read all six type variations, but no two children in each of the two large groups of 36 pupils took the same combinations of test form, type variation, and test order. The results were analyzed for the significance of the statistical variation induced by each of the components analyzed, with particular attention to the significance of the variance assignable to type variation. The *F*-value for type variation was insignificant at the .05 level for both criteria: reading speed and eye-blink frequency. For the speed scores, the variance attributable to individual differences, test form, and order of testing were all highly significant. Only variance induced by individual differences was significant for the eye-blink scores.

The scores of a group of 40 myopes and a group of 10 hyperopes were then analyzed to ascertain whether there was a difference between the means of the type size scores, a difference which might be attributed to the type of visual impairment. For each of these two groups the differences between the mean scores of the various type sizes were tested for significance. The results were such as to indicate that no real differences existed.

An effort was also made to test the degree of relationship between the two criteria: reading speed and eye-blink frequency. The results indicated that practically no relationship obtains in this respect with the group of subjects studied.

V

ANALYSIS AND INTERPRETATION OF THE RESULTS

The results of the statistical analysis of the data may be summarized as follows:

1. Analysis of variance revealed no significant differences in speed with which the different type variations were read.
2. Analysis of variance revealed no significant differences in the frequency of eye-blink for each of the type variations.
3. The speed criterion revealed significantly large differences among the test forms and the factor of testing order. Eye-blink scores did not reflect these differences.
4. There was no statistically reliable difference among the means of the type variation scores in the case of either the myopic or the hyperopic subjects. Nevertheless, suggestively patterned differences between these groups seemed to favor the larger types for the hyperopes.
5. There appeared to be no relationship between the criteria employed.
6. For adult observers of normal or near-normal vision, 24-point type is the most visible of the types studied (as measured by the Luckiesh-Moss Visibility Meter), followed in order by: Ampli-type, typed (dark); 18-point; Ampli-type, typed (light); Ampli-type, Mimeographed; 14-point; and 12-point.

POSSIBLE INTERPRETATIONS OF THE RESULTS

The complete absence of significant differences among the type variations with respect to reading speed and eye-blink

frequency is, at first, rather surprising. On the basis of previous studies and common observation, we should be led to expect a real difference between at least the extremes of 12-point and 24-point type. The absence of positive results is subject to at least three interpretations:

1. For visually handicapped children, reading for brief periods, there are no real differences in the readability of the type variations studied.
2. For visually handicapped children there are real differences in the readability of the type variations studied, but the criteria used did not measure them adequately.
3. For visually handicapped children there are real differences in the readability of the type variations studied, but they were obscured by error for which the experimental design provided no adequate estimate, and which swelled the error variance.

EVALUATION OF POSSIBLE INTERPRETATIONS

It is impossible, of course, to say which of these three interpretations of the data is the correct one. Perhaps the absence of positive results cannot be attributed to any one factor alone. Nevertheless, the data should be examined to determine whether one interpretation seems more plausible than another.

Interpretation 1 runs counter to popular belief for a number of reasons. For example, there are many investigations reporting differences in the readability of less widely differing type sizes than those studied in this experiment. It is true that almost all such studies are concerned with smaller type sizes and with subjects of normal vision; hence, they are really not comparable with this investigation. Nevertheless, they do predispose one to expect differences in readability among types differing as much as those examined in the present study.

The fact that school children show a strong preference for books with relatively larger types lends credence to this viewpoint. It can also be pointed out that the practice of using large

type in the sight-saving classes is based on more than Irwin's findings.[1] The teachers of these classes relate that type sizes less than 18 points in size cause a number of pupils to squint and assume unnatural reading positions, presumably because such types are difficult for them to read. Some pupils actually complain of discomfort from reading types smaller than 18 points in size. The visibility ratings in Table 1 (p. 27) also support this belief.

Nevertheless, the results of this study are such as to render interpretation 1 a real possibility. In the variances of the reading speed scores, for example, it will be noted that the statistical tool was precise enough, and the criterion sufficiently sensitive, to reveal significantly large variances for test form and order of testing. It is reasonable to assume that had there existed comparable differences with respect to type variation they would have been made evident. Apparently, therefore, for the subjects tested in this study, there is not as great a difference in the readability of the types investigated as seems to be commonly assumed —at least when reading for brief periods of time.

Validity and reliability of the criteria

Turning our attention to interpretation 2, we are faced with the question of the adequacy of the criteria. In Chapter II it was indicated that Luckiesh and Moss consider reading speed to be too insensitive to changes in the severity of the visual task to be of much value as a criterion, particularly in the brief span of 5 minutes.[2] On the other hand, the criterion of eye-blink frequency, endorsed virtually without qualification by Luckiesh and Moss, has not, to the knowledge of the writer, been validated by any investigators except these two. This takes us back to the problem at the beginning of Chapter II; namely, that there

[1] See page 3 for Irwin's conclusions.

[2] M. Luckiesh and F. Moss, *Visibility and Readability of Print on White and Tinted Papers,* p. 3. Publication 271, National Society for the Prevention of Blindness, Inc., New York. (Reprinted from *The Sight-Saving Review,* Vol. 8, No. 2, June, 1938.)

is no criterion of readability, or of visual effort or fatigue, which is universally agreed upon as being valid and reliable.

The validity of the criterion of reading speed is, as far as the writer is aware, unquestioned; its severest critics grant that it may measure (albeit insensitively) readability. The fact that it is insensitive to relatively small changes in readability is its greatest drawback. That is admitted by its warmest proponents. Its validity, however, has been attested to by the number of investigators who have reported it to reflect differences in readability, when the conditions were such as to make such differences axiomatic.

Eye-blink frequency, on the other hand, has not been as widely validated. One authority, indeed, questions the validity of the criterion on the grounds that Luckiesh and Moss misinterpret the Ponder and Kennedy conclusions on muscle tension.[3] Whereas there seems to be no reported evidence disputing this criterion, there is also no confirming evidence from sources other than the Luckiesh and Moss reports. It is notable that in the present study eye-blink frequency was revealed to bear little relationship to reading speed (the latter accepted as a valid criterion of readability) when measured over periods as brief as five minutes. It may be, therefore, that there are real differences in the readability of some of the types studied, but that the speed criterion was too crude to reveal them, and that the eye-blink criterion was not a valid measure of them.

Possible effects of undetermined systematic error

As analysis of variance is probably the best statistical method yet devised for estimating the amount of error in an experiment of this kind, interpretation 3 would seem unlikely. There are, however, sources of error which were not quantitatively isolated in the experiment, and these may have swelled the residual variance (used in obtaining the F-value) to such an extent that it

[3] M. A. Tinker, "Illumination Intensities for Reading," *American Journal of Ophthalmology*, 18:1036–38, 1935.

obscured any real differences in the data. One of these sources is the possible unreliability of a five-minute sampling period. Whereas other investigators have reported positive results in experiments using even shorter sampling periods, their other conditions were very different from those of this experiment (adult subjects, normal vision, etc.). It is possible, therefore, that the sampling periods were not long enough to provide adequate reliability.

Another possible source of error is the variation within and among the observers who counted eye-blink—the "unreliability" of the observers, so to speak. Some observers consistently found lower eye-blink totals than others, as is revealed in Table 9. Unfortunately, this source of error was not randomized, as its magnitude was unforeseen. The differences among the observers in this respect, however, does suggest that eye-blink frequency count may not be an objective enough measure of visual effort expended when human observation is the means of determining it. Furthermore, "eye-blink" itself is in need of definition, and even when defined is necessarily individually interpreted. What *is* an eye-blink? Must the upper lid come in contact with the lower lid? If not, how far down *must* it come to be counted as a definite blink? Half-way? Three-fourths of the way? Whatever standard is set, it is a practical impossibility for the unassisted human visual mechanism to determine whether

TABLE 9

NUMBER OF EYE-BLINKS COUNTED BY EACH OBSERVER, WITH MEANS

Observer	*Total Eye-Blinks Counted*	*Mean*
1.	784	10.88
2.	823	11.43
3.	1203	16.71
4.	1336	18.56
5.	1278	17.75
6.	1396	19.39

the lids did come in contact with each other, or whether the upper lid came half-way down, or only seven-sixteenths of the way. There is, therefore, a real possibility that this criterion is unreliable when the count is made by human beings unassisted by mechanical devices. An experiment in which pairs of investigators simultaneously counted the blinks of the same subjects might be revealing in this respect.

One other result may be interpreted interestingly. When scores of hyperopes and myopes were separated out, and their means tested for significant differences, the means of the eye-blink scores (Table 7, p. 50) seemed to suggest that the hyperopes blinked appreciably more than the myopes on the 12- and 14-point types, and appreciably less on the larger types. The greatest differences in this latter respect occurred in the case of typing and 24-point Caslon. These differences suggest that the larger types may be more readable for the hyperopes. Less strongly, they point to the possibility of the smaller types being the more readable for the myopes. The differences are not large enough, however, to approach conclusiveness.

As has been pointed out, there was a heavy preponderance of myopes in the experimental population, and relatively few hyperopes. It is entirely possible that, if a large enough group of hyperopes were included in a study of this kind, it might be found that for them there are real differences in the readability of the experimental types. It is also possible that real differences between them and the myopes, with respect to the readability of the experimental types, might be revealed.

Thus, it is evident that there is a possibility that uncontrolled factors operated to obscure what differences the criteria might otherwise have revealed. Nevertheless, this does not seem to be a strong possibility in the face of the fact that: (1) a large part of the total variance was accounted for by other systematic factors; (2) the residual (error) variance was relatively small; and (3) the variance ascribable to type form was even less than that predicted by the error variance.

VI

SUMMARY AND CONCLUSIONS

THE PROBLEM

There is a need for objective evidence on the readability of various type sizes for children whose vision is severely impaired. This study was undertaken in an attempt to supply data on the relative readability of several type sizes and variations. For the purposes of this study, readability is defined as the extent to which a given type size or form lends itself to being read with absence of visual effort—with least cost to the human visual mechanism—when comprehension is held relatively constant. The types studied were:

1. 12-point Caslon No. 3, Linotype
2. 14-point Caslon No. 3, Linotype
3. 18-point Caslon Bold No. 79, Monotype
4. 24-point Caslon Bold No. 79, Monotype
5. Ampli-Type, typed
6. Ampli-Type, Mimeographed

THE CRITERIA

The criteria employed to measure readability were two: frequency of involuntary eye-blink (which has been used extensively, and almost exclusively, in the readability studies of Luckiesh and Moss), and speed of reading. The relative visibility of each of the types studied was also measured, using the Luckiesh-Moss Visibility Meter.

THE TESTS

The reading tests used were made up of items from the Gates speed of reading tests of the Gates Reading Survey and the

Modern School Achievement Test, plus 72 items, closely similar to the items in the above-mentioned tests, written by the author of the study. The entire 300 items were then randomly distributed so as to make up six test forms of 50 items each. The tests are of about third grade difficulty level.

Each test form was printed in all six of the type variations under study. As there were six test forms and six type variations, there resulted thirty-six different combinations of test form and type variation.

The typographical conditions of the tests were determined by the advice of several experts in the fields of typography and book designing (see pages 24 to 28). All tests were printed on paper 8.5 inches by 11.0 inches.

PAPER

The paper upon which the tests were printed was Warren's Olde Style Wove, India, Substance 60. It is a light cream-buff colored book paper of high quality, with a dull, light-diffusing surface. Its characteristics with respect to glare, opacity, smoothness, and color were ascertained quantitatively, and are specified on pages 29–30.

THE SUBJECTS

Seventy-two pupils from the sight-saving classes of New York City public schools were the subjects. They ranged in age from 9 to 14 years, and in school grade placement from the fourth to eighth year. The great majority were from 10 to 12 years old, and were in grades 5 and 6. Twenty-nine were boys, and forty-three were girls.

Of the total, fifty-two were myopic, ten were hyperopic, and the other ten were distributed over seven other types of visual anomaly. Many of the myopic and hyperopic cases were also afflicted with astigmatism, and a few had other complications.

According to the judgment of the classroom teachers of the

subjects, all the pupils tested had reading abilities above the third grade level; the great majority were judged to be of good fifth grade level in reading.

ADMINISTRATION OF THE TESTS

For the test administration, the pupils were divided into twelve groups of six pupils each. The tests were administered in a Latin Square pattern in such a manner that each pupil took six test forms, each printed in a different type variation, but no two children in each group of thirty-six took the same six combinations in the same order.

The tests were administered to six pupils at a time, with no other pupils present. An adult observer (one for each subject), seated beside and slightly behind the subject, counted the number of eye-blinks during each five minute period in which a test was administered. During the test administration each desk was receiving from 27 to 35 foot-candles of light (determined with a General Electric Light Meter). At the end of each five-minute test period, the number of eye-blinks was noted on the title page of the test by the observer. There were three-minute rest periods between all but the third and fourth tests, and between these two was a five-minute relaxation period. The eye-blinks were observed and counted mentally.

METHODS OF ANALYSIS OF THE DATA

The variance of the data was analyzed by means of a five-component analysis of variance. The sources of variance which were isolated were: variance among individuals; variance among test forms; variance arising from test order; variance attributable to differences in type; and residual variance. This analysis was made for each of the two criteria, reading speed and eye-blink frequency. The *F*-test of significance was employed to evaluate the degree to which each systematic variance was significantly larger than the error variance.

In another treatment of the data, two groups were selected

from the total number of subjects: a group of forty myopes and a group of ten hyperopes. Within each of these two groups the means of the criterion scores for each type variation were computed, and the significance of the difference between these means was ascertained. This analysis was made to determine whether there was a statistically reliable difference among the means of the criterion scores for myopic and hyperopic children, a difference which might have been obscured when the data of both groups were lumped together in one treatment.

The data were also examined to discover whether there was apparent a relationship between the criteria (reading speed and eye-blink frequency), which we should expect if they both measure the same thing. For each subject, the relationship between the scores for each of the criteria was ascertained, and the number of times in which there was a direct relationship, an inverse relationship, and no relationship was determined.

The type variations were also tested for visibility by adult observers of normal or near-normal vision, using the Luckiesh-Moss Visibility Meter.

RESULTS

The results of the foregoing analyses of the data were as follows:

1. The analysis of variance revealed no significant differences in the speed with which the different type variations were read.
2. The analysis of variance revealed no significant differences in the frequency of eye-blink for each of the type variations.
3. The speed criterion revealed significantly large differences among the test forms and the position in order of testing which did not show up in the eye-blink scores.
4. There was no statistically reliable difference among the means of the type variation scores in the case of either the myopic or the hyperopic subjects. Nevertheless, sugges-

tively patterned differences between these groups seemed to favor the larger types for the hyperopes.

5. There appeared to be no relationship between the criteria employed.
6. For adult observers of normal or near-normal vision, 24-point type is the most visible of the types studied (as measured by the Luckiesh-Moss Visibility Meter), followed in order by: Ampli-type, typed (dark); 18-point; Ampli-type, typed (light); Ampli-type, Mimeographed; 14-point; and 12-point.

INTERPRETATION OF THE RESULTS

Three general interpretations of the results seem to be possible:

1. For visually handicapped children, reading for brief periods, there are no real differences in the readability of the types studied.
2. For visually handicapped children there are real differences in the readability of the type variations studied, but the criteria used did not measure them adequately.
3. For visually handicapped children there are real differences in the readability of the type variations studied, but they were obscured by error for which the experimental design provided no adequate estimate, and which swelled the error variance.

Examination of the results leads to the belief that interpretation 3 is the least likely. On the other hand, there is evidence that the criteria—particularly the criterion of eye-blink frequency—did not measure adequately what differences there may have been. This may have been because of the brevity of the sampling periods. A strict interpretation of the results of the statistical analysis would incline one to state that, for the visually handicapped children studied, there is no real difference in the readability of the type variations tested. In spite of the lack of statistically significant differences, however, there is a

possibility that for the subjects used in this study there are differences in the readability of the type forms investigated but they were obscured by the operation of some other factor or combination of factors.

CONCLUSIONS

The results of this study will not permit the conclusion that any one of the experimental types is preferable to any of the others for use in sight-saving classes. Certain other conclusions seem tenable, however:

1. For the subjects used in this study, reading for five-minute periods under from 27 to 35 foot-candles of illumination, there is not as great a difference in the readability of the type forms studied as has been commonly supposed.
2. There is a need for further evaluation of the validity and reliability of eye-blink frequency as a criterion of readability.[1]
3. Studies of this type should employ sampling periods longer than five minutes, or space a number of brief samples from a longer total working period.
4. Studies of readability (as herein defined) cannot be conclusive until and unless some more valid and sensitive criterion of visual fatigue, or of visual effort expended, is developed.

[1] Recent studies conducted at Harvard University yield additional evidence to support the belief that eye-blink frequency may be a poor criterion for the measurement of visual effort expended, or of visual fatigue. The report of these studies was received too late to be incorporated into the body of this manuscript. *See* R. A. McFarland, A. H. Holway, and L. M. Hurvich, *Studies of Visual Fatigue*, pp. 71–86. Harvard University Press, Cambridge, Mass., 1942.

BIBLIOGRAPHY

ALDERMAN, EVERETT. "The Effect of Size of Type on Speed of Reading." *Pittsburgh Schools,* Vol. 13, No. 2, November–December, 1938.

BAIRD, J. W. "The Legibility of a Telephone Directory." *Journal of Applied Psychology,* Vol. 1, pp. 30–37, 1917.

BERENS, CONRAD AND HATHAWAY, WINIFRED. "Sight-Saving Classes." *American Journal of Ophthalmology,* Vol, 18, pp. 845–849, 1935.

BETTS, EMMETT A. *The Prevention and Correction of Reading Difficulties.* Chapter IX, "The Mechanics of Reading Materials." Row, Peterson and Co., New York, 1937.

BLACKHURST, J. H. *Investigations in the Hygiene of Reading.* Warwick and York, Inc., Baltimore, 1927.

BUCKINGHAM, B. R. "New Data on the Typography of Textbooks." *The Textbook in American Education: Thirtieth Yearbook of the National Society for the Study of Education,* Part II, pp. 93–95. Public School Publishing Co., Bloomington, Ill., 1931.

BUCKINGHAM, B. R. "Type Sizes of First-Grade Textbooks Are Tested for Readability." *Nations Schools,* Vol. 12, pp. 45–48, 1933.

BURTT, H. E. AND BASCH, C. "Legibility of Bodoni, Baskerville Roman, and Cheltenham Type Faces." *Journal of Applied Psychology,* Vol. 7, pp. 237–245, 1923.

COMMITTEE FOR THE STUDY OF THE CARE AND EDUCATION OF PHYSICALLY HANDICAPPED CHILDREN IN THE PUBLIC SCHOOLS OF THE CITY OF NEW YORK. *Report of the Reading Sub-Committee for the Visually Handicapped.* Board of Education of the City of New York, 1940.

FEREE, C. E. AND RAND, G. "Instrument for Measuring Dynamic Speed of Vision, Speed of Accommodation, and Ocular Fatigue." *Archives of Ophthalmology,* Vol. 15, pp. 1072–1087, June, 1936.

GAGE, HARRY L. "Research in Readability, IV: Comparison of Caslon Boldface, Textype, and Memphis Medium. Analysis of the Results of Luckiesh and Moss." *Linotype News,* Vol. XVII, No. 2, July, 1938.

GATES, ARTHUR I. *Gates Reading Survey for Grades 3–10,* Forms I and II. Bureau of Publications, Teachers College, Columbia University, New York, 1939.

GATES, ARTHUR I. *Manual of Directions for the Gates Reading Survey.* Bureau of Publications, Teachers College, Columbia University, New York.

GATES, ARTHUR I. *Manual of Directions for the Modern School Achievement Test.* Bureau of Publications, Teachers College, Columbia University, New York.

GATES, ARTHUR I. *Modern School Achievement Test, Short Form* (Speed of Reading), Forms I and II. Bureau of Publications, Teachers College, Columbia University, New York.

GATES, ARTHUR I. "What Do We Know about Optimum Lengths of Line in Reading?" *Journal of Educational Research,* Vol. XXIII, pp. 1–7, January, 1931.

GILLILAND, A. R. "The Effect on Reading of Changes in the Size of Type." *Elementary School Journal,* Vol. 24, pp. 138–146, 1923.

GREENE, E. B. "The Legibility of Typewritten Material." *Journal of Applied Psychology,* Vol. 17, pp. 713–728, 1933.

GREENE, E. B. "The Relative Legibility of Linotyped and Typewritten Material." *Journal of Applied Psychology,* Vol. 18, pp. 697–704, 1934.

GRIFFING, H. AND FRANZ, S. I. "On the Conditions of Fatigue in Reading." *Psychological Review,* Vol. 3, pp. 513–530, 1896.

HOLMES, G. "The Relative Legibility of Black Print and White Print." *Journal of Applied Psychology,* Vol. 15, pp. 248–251, 1931.

HOVDE, H. F. "The Relative Effects of Size of Type, Leading, and Context." *The Journal of Applied Psychology,* Vol. 13, pp. 600–629, 1929 (Part I) and Vol. 14, pp. 63–73, 1930 (Part II).

HUEY, E. B. *The Psychology and Pedagogy of Reading.* The Macmillan Co., New York, 1909.

IRWIN, ROBERT B. "Large Type Reading Tests" (Unpublished report). American Foundation for the Blind, Inc., New York.

KURTZ, J. I. "An Experimental Study of Ocular Fatigue." *American Journal of Optometry,* Vol. 15, pp. 86–117, 1938.

LINDQUIST, E. F. *A First Course in Statistics.* Houghton Mifflin Company, Boston, 1938.

LINDQUIST, E. F. *Statistical Analysis in Educational Research.* Houghton Mifflin Company, Boston, 1940.

LUCKIESH, M. AND MOSS, F. K. "A Summary of Researches Involving Blink-Rate as a Criterion of Ease of Seeing." *Illuminating Engineering,* Vol. 35, pp. 19–32, January, 1940.

LUCKIESH, M. AND MOSS, F. K. "Boldness as a Factor in Type Design and Typography." *Journal of Applied Psychology,* Vol. 24, pp. 170–183, April, 1940.

LUCKIESH, M. AND MOSS, F. K. "Criteria of Readability." *Journal of Experimental Psychology,* Vol. 27, pp. 256–270, 1940.

LUCKIESH, M. AND MOSS, F. K. "Effects of Classroom Lighting Upon Educational Progress and Visual Welfare of School Children." *Illuminating Engineering,* Vol. 35, pp. 915–938, 1940.

LUCKIESH, M. AND MOSS, F. K. "Effects of Leading on Readability." *Journal of Applied Psychology,* Vol. 22, pp. 140–160, April, 1938.

Luckiesh, M. and Moss, F. K. "The Eyelid Reflex as a Criterion of Ocular Fatigue." *Journal of Experimental Psychology,* Vol. 20, pp. 589–596, 1937.

Luckiesh, M. and Moss, F. K. "Fatigue of Convergence Induced by Reading as a Function of Illumination Intensity." *American Journal of Ophthalmology,* Vol. 18, pp. 319–323, April, 1935.

Luckiesh, M. and Moss, F. K. "Frequency of Blinking as a Clinical Criterion of Ease of Seeing." *American Journal of Ophthalmology,* Vol. 22, pp. 616–621, 1939.

Luckiesh, M. and Moss, F. K. "The Quantitative Relationship Between Visibility and Type Size." *Journal of the Franklin Institute,* Vol. 227, pp. 87–97, January, 1939.

Luckiesh, M. and Moss, F. K. "The Readability of Stencil Duplicated Materials." *Sight-Saving Review,* Vol. 9, pp. 295–304, 1939.

Luckiesh, M. and Moss, F. K. "The Relative Visibility of Print in Terms of Illumination Intensity." *Sight-Saving Review,* Vol. 5, pp. 272–280, December, 1935.

Luckiesh, M. and Moss, F. K. *The Science of Seeing.* D. Van Nostrand Company, Inc., New York, 1937.

Luckiesh, M. and Moss, F. K. "Size of Pupil as a Possible Index for Visual Fatigue." *American Journal of Ophthalmology,* Vol. 16, pp. 393–396, May, 1933.

Luckiesh, M. and Moss, F. K. "Some Effects of Illumination on Reading Efficiency." *Illuminating Engineering,* Vol. 35, pp. 703–707, September, 1940.

Luckiesh, M. and Moss, F. K. "Supra-Threshold Visibility." *Journal of the Optical Society,* Vol. 30, pp. 62–69, February, 1940.

Luckiesh, M. and Moss, F. K. *Type Visibility: Sixth Annual Production Yearbook.* Colton Press, Inc., New York, 1940.

Luckiesh, M. and Moss, F. K. "Visibility and Ease of Seeing." *Industrial Medicine,* p. 33, April, 1940.

Luckiesh, M. and Moss, F. K. "The Visibility and Readability of Printed Matter." *Journal of Applied Psychology,* Vol. 23, pp. 645–659, 1939.

Luckiesh, M. and Moss, F. K. *Visibility and Readability of Print on White and Tinted Papers.* Publication 271, National Society for the Prevention of Blindness, Inc. New York. Reprinted from *Sight-Saving Review,* Vol. 8, pp. 123–134, June, 1938.

Luckiesh, M. and Moss, F. K. "Visibility: Its Measurement and Significance in Seeing." *Journal of the Franklin Institute,* Vol. 220, p. 431, 1935.

Luckiesh, M. and Moss, F. K. "Visibility of Various Type Faces." *Journal of the Franklin Institute.* Vol. 223, p. 77, 1937.

McFarland, R. A., Knehr, C., and Berens, C. "Metabolism and Pulse

Rate as Related to Reading under High and Low Levels of Illumination." *Journal of Experimental Psychology*, Vol. 25, pp. 65–75, July, 1939.

McFarland, R. A., Holway, A. H., and Hurvich, L. M. *Studies of Visual Fatigue*, pp. 71–86. Harvard University Press, Cambridge, Mass., 1942.

Miyake, R., Dunlap, J. W., and Cureton, E. E. "The Comparative Legibility of Black and Colored Numbers on Colored and Black Backgrounds." *Journal of General Psychology*, Vol. 3, pp. 340–343, 1930.

Ponder, E. and Kennedy, W. P. "On the Act of Blinking." *Journal of Experimental Physiology*, Vol. 18, pp. 89–110, 1927.

Robertson, C. J. "Effect of Fatigue on the Adjustment of the Eye to Near and Far Vision." *Archives of Ophthalmology*, Vol. 17, pp. 859–876, 1937.

Roethlein, B. E. "The Relative Legibility of Different Faces of Printing Types." *American Journal of Psychology*, Vol. 23, pp. 1–36, 1912.

Russell, David W. "Should Standards Be Adopted for Stencil Duplicated Classroom Materials?" *American School Board Journal*, Vol. 99, pp. 52–53 and 92, July, 1939.

Sanford, E. C. "The Relative Legibility of Small Letters." *American Journal of Psychology*, Vol. 1, pp. 402–435, 1888.

Snedecor, G. W. *Statistical Methods*. Collegiate Press, Inc., Ames, Iowa, 1938.

Taylor, C. D. "The Relative Legibility of Black and White Print." *Journal of Educational Psychology*, Vol. 25, pp. 561–578, 1934.

Telford, C. W. and Thompson, N. "Some Factors Influencing Voluntary and Reflex Eyelid Responses." *Journal of Experimental Psychology*, Vol. 16, pp. 524–539, 1933.

Tinker, M. A. "Illumination Intensities for Reading." *American Journal of Ophthalmology*, Vol. 18, pp. 1036–1038, 1935.

Tinker, M. A. "The Relative Legibility of the Letters, the Digits, and of Certain Mathematical Signs." *Journal of General Psychology*, Vol. 1, pp. 472–496, 1928.

Tinker, M. A. and Paterson, D. G. "Studies of Typographical Factors Influencing Speed of Reading: II, Size of Type." *Journal of Applied Psychology*, Vol. 13, pp. 120–130, 1929.

Tinker, M. A. and Paterson, D. G. "Studies of Typographical Factors Influencing Speed of Reading: V, Simultaneous Variation of Type Size and Line Length." *Journal of Applied Psychology*, Vol. 15, pp. 72–78, 1931.

Tinker, M. A. and Paterson, D. G. "Studies of Typographical Factors Influencing Speed of Reading: IX, Reductions of Size of Newspaper Print." *Journal of Applied Psychology*, Vol. 16, pp. 525–531, October, 1932.

Tinker, M. A. and Paterson, D. G. "Studies of Typographical Factors Influencing Speed of Reading: X, Style of Type Face." *Journal of Applied Psychology,* Vol. 16, pp. 605–613, 1932.

Tinker, M. A. and Paterson, D. G. "Studies of Typographical Factors Influencing Speed of Reading: XI, Role of Set in Typographical Studies." *Journal of Applied Psychology,* Vol. 19, pp. 647–651, 1935.

Tinker, M. A. and Paterson, D. G. "Studies of Typographical Factors Influencing Speed of Reading: XII, Printing Surface." *Journal of Applied Psychology,* Vol. 20, pp. 128–131, 1936.

Volk, Kurt. *Using Type Correctly.* Kurt H. Volk, Inc., New York, 1935.

Wager, R. E. "A Method of Measuring Fatigue of the Eyes." *Journal of Educational Psychology,* Vol. 13, pp. 561–570, 1922.

Webster, H. A. and Tinker, M. A. "The Influence of Type Face on the Legibility of Print." *Journal of Applied Psychology,* Vol. 19, pp. 43–52, 1935.

Webster, H. A. and Tinker, M. A. "The Influence of Paper Surface on the Perceptibility of Print." *Journal of Applied Psychology,* Vol. 19, pp. 145–147, 1935.

Weiss, A. P. "The Focal Variator." *Journal of Experimental Psychology,* Vol. 2, pp. 106–113, 1917.

White House Conference on Child Health and Protection. *Special Education: The Handicapped and the Gifted.* The Century Company, New York, 1931.

c. 1
HV1737 McNally, Harold Joseph.
M232 THE READABILITY OF CERTAIN
TYPE SIZES AND FORMS.... (1943)

Date Due

HV1737 c.1
M232
McNally, Harold Joseph.
AUTHOR
THE READABILITY OF CERTAIN TYPE
TITLE SIZES AND FORMS IN SIGHT-
SAVING CLASSES. (1943)

DATE DUE	BORROWER'S NAME
Reference Copy	